THE UNIVERSITY
of EDINBURGH

PHENOME WIDE ASSOCIATION STUDY OF VITAMIN D GENETIC VARIANTS IN THE UK BIOBANK COHORT

By
Xiangrui Meng

Doctor of Philosophy
The University of Edinburgh
August 2018

DECLARATION

Thesis: Phenome Wide Association Study of Vitamin D Genetic Variants in the UK Biobank Cohort

I, Xiangrui Meng hereby declare that I am the sole author of this thesis. I developed the hypotheses examined in this thesis and conducted all aspects of the research except for systematic reviews, which were conducted with the assistance from my colleagues, Xue Li and Yazhou He. This thesis has not been submitted for any other degree or professional qualification.

Signature:  **Date:** 2018/10/24

AKNOWLEDGMENTS

I would like to thank my principal supervisor **Dr. Evropi Theodoratou** and second supervisor **Professor Harry Campbell**, for their guidance and encouragement, throughout my PhD.

I would like to thank all **Participants in the UK Biobank Cohort** for their priceless participation, which would change every aspect of health care system and industry for the coming generations.

I would like to thank **Professor Haijun Wang**, my supervisor of MSc Degree back in Peking University, for her great help of bringing me into academia.

I wish to acknowledge **Dr. Maria Timofeeva**, **Dr. Athina Spiliopoulou** and **Dr. Peter Joshi**, for their kindly help in genetic and statistical analysis methods.

I wish to acknowledge **Dr. Lina Zgaga** and **Professor Jim Wilson**, for their attendance in the first and second year review of my project and all the wonderful suggestions they gave.

I wish to acknowledge **Professor Denny Joshua** and his colleagues in Vanderbilt University, USA, for the good collaboration job they have done in analysis package.

I wish to acknowledge the **China Scholarship Council** in China, for funding me and giving me the opportunity to conduct this thesis.

Finally, I would like to thank my parents and all of my friends, for their support and encouragement through my PhD.

LAY SUMMARY OF THESIS

The link between vitamin D concentrations and various conditions and diseases has been assessed in a large and rapidly expanding literature. Although vitamin D can be synthesized in the skin with sun exposure, vitamin D deficiency is of high prevalence worldwide. It is estimated that around 1 billion people worldwide suffer from vitamin D deficiency or insufficiency. This led to intensive research activity in relation to vitamin D and multiple health outcomes.

Existing evidence from previous studies is largely inconclusive. Therefore, in this thesis, I would like to explore the link between genetically determined vitamin D and the whole spectrum of health outcomes using a large study, called the UK Biobank.

In this study, I first studied whether 6 genetic variants, which were found to be relevant to vitamin D level, were linked to all health outcomes. Then I explored whether vitamin D level could impact nine health outcomes biologically, including systolic blood pressure, diastolic blood pressure, body mass index, risk of hypertension, type 2 diabetes, ischemic heart disease, depression, non-vertebral fracture and all-cause mortality. Given the large sample size of UK Biobank we would be able to observe links between vitamin D and common diseases. When individual genetic mutations affecting vitamin D were tested, only one was found to be linked to calculus outcomes. Beyond that, in this study I did not find any moderate to large effect of vitamin D on any of the nine outcomes. Thus, my study did not support the hypothesis that vitamin D plays a role in any health outcomes. Further studies of even larger size are needed to explore whether smaller effects of vitamin D exist, which could not be detected by my study.

ABSTRACT

Introduction

Vitamin D status is an important public health issue due to the high prevalence of vitamin D insufficiency and deficiency, especially in high latitude areas. Furthermore, it has been reported to be associated with a number of diseases. In a previous umbrella review of meta-analyses of randomized clinical trials (RCTs) and of observational studies, it was found that plasma/ serum 25-hydroxyvitamin D (25(OH)D) or supplemental vitamin D has been linked to more than 130 unique health outcomes. However, the majority of the studies yielded conflicting results and no association was convincing.

Aim and Objectives

The aim of my PhD was to comprehensively explore the association between vitamin D and multiple outcomes. The specific objectives were to: 1) update the umbrella review of meta-analysis of observational studies or randomized controlled trials on associations between vitamin D and health outcomes published between 2014 and 2018; 2) conduct a systematic literature review of previous Mendelian Randomization studies on causal associations between vitamin D and all outcomes; 3) conduct a systematic literature review of published phenome wide association studies, summarizing the methods, results and predictors; 4) create a polygenic risk score of vitamin D related genetic variants, weighted by their effect estimates from the most recent genome wide association study; 5) encode phenotype groups based on electronic medical records of participants; 6) study the associations between vitamin D related SNPs and the whole spectrum of health outcomes, defined by electronic medical records utilising the UK Biobank study; 7) explore the causal effect of 25-hydroxyvitamin D level on health outcomes by applying novel instrumental variable methods.

Methods

First I updated the vitamin D umbrella review published in 2015, by summarizing the evidence from meta-analyses of observational studies and meta-analyses of RCTs published between 2014 and 2018. I also performed a systematic literature review of

all previous Mendelian Randomizations studies on the effect of vitamin D on all health outcomes, as well as a systematic review of all published PheWAS studies and the methodology they applied. Then I conducted original data analysis in a large prospective population-based cohort, the UK Biobank, which includes more than 500,000 participants. A 25(OH)D genetic risk score (weighted sum score of 6 serum 25(OH)D-related SNPs: rs3755967, rs12785878, rs10741657, rs17216707, rs10745742 and rs8018720, as identified by the largest genome wide association study of 25(OH)D levels) was constructed to be used as the instrumental variable. I used a phenotyping algorithm to code the electronic medical records (EMR) of UK Biobank participants into 1853 distinct disease categories and I then ran the PheWAS analysis to test the associations between the 25(OH)D genetic risk score and 950 disease outcome groups (i.e. outcomes with more than 200 cases). For phenotypes found to show a statistically significant association with 25(OH)D levels in the PheWAS or phenotypes which were found to be convincing or highly suggestive in previous studies, I developed an extended case definition by incorporating self-reported data collected by UK Biobank baseline questionnaire and interview. The possible causal effect of vitamin D on those outcomes was then explored by the MR two-stage method, inverse variance weighted MR and Egger's regression, followed by sensitivity analyses.

Results

In the updated systematic literature review of meta-analyses of observational studies or RCTs, only studies on new outcomes which had not been covered by the previous umbrella review were included. A total of 95 meta-analyses met the inclusion criteria. Among the included studies there were 66 meta-analyses of observational studies, and 29 meta-analyses of RCTs. Eighty-five new outcomes were explored by meta-analyses of observational studies, and 59 new outcomes were covered by meta-analyses of RCTs.

In the systematic review of published Mendelian Randomization studies on vitamin D, a total of 29 studies were included. A causal role of 25(OH)D level was supported by MR analysis for the following outcomes: type 2 diabetes, total adiponectin, diastolic

blood pressure, risk of hypertension, multiple sclerosis, Alzheimer's disease, all-cause mortality, cancer mortality, mortality excluding cancer and cardiovascular events, ovarian cancer, HDL-cholesterol, triglycerides and cognitive functions.

For the systematic literature review of published PheWAS studies and their methodology, a total of 45 studies were included. The processes for implementing a PheWAS study include the following steps: sample selection, predictor selection, phenotyping, statistical analysis and result interpretation. One of the main challenges is the definitions of the phenotypes (i.e., the method of binning participants into different phenotype groups). In the phenotyping step, an ICD curated phenotyping was widely used by previous PheWAS, which I also used in my own analysis.
By applying the ICD curated phenotyping, 1853 phenotype groups were defined in the participants I used. In PheWAS, only phenotype groups with more than 200 cases were analysed (920 phenotypes). In the PheWAS, only associations between rs17216707 (*CYP24A1*) and "calculus of ureter" (beta = -0.219, se = 0.045, $P = 1.14*10^{-6}$), "urinary calculus" (beta = -0.129, se = 0.027, $P = 1.31*10^{-6}$), "alveolar and parietoalveolar pneumonopathy" (beta = 0.418, se = 0.101, $P = 3.53*10^{-5}$) survived Bonferroni correction.

Nine outcomes, including systolic blood pressure, diastolic blood pressure, body mass index, risk of hypertension, type 2 diabetes, ischemic heart disease, depression, non-vertebral fracture and all-cause mortality were explored in MR analyses. The MR analysis had more than 80% power for detecting a true odds ratio of 1.2 or larger for binary outcomes. None of explored outcomes were statistically significant. Results from multiple MR methods and sensitivity analyses were consistent.

Discussion

Vitamin D and its association with multiple outcomes has been widely studied. More than 230 outcomes have been linked with vitamin D by meta-analyses of observational studies and RCTs. On the contrary, evidence from Mendelian Randomization studies is lacking. In particular I identified only 20 existing MR studies and only 13 outcomes were suggested to be causally related to vitamin D. In the systematic literature review

of previous PheWAS studies, I summarized the applied methods, predictors and results. Although phenotyping based on ICD codes provided good performance and was widely applied by previous PheWAS studies, phenotyping can be improved if lab data, imaging data and medical notes can be incorporated. Alternative algorithms, which takes advantage of deep learning and thus enable high precision phenotyping, needs to be developed.

From the PheWAS analysis, the score of vitamin D related genetic variants was not statistically significantly associated with any of the 920 phenotypes tested. In the single variant analysis, only rs17216707 (*CYP24A1*) was shown to be associated with calculus outcomes statistically significantly. Previous studies reported associations between vitamin D and hypercalcemia, hypercalciuria, nephrolithiasis and nephrocalcinosis, may be due to the role of vitamin D in calcium homeostasis.

In the MR analysis, I found no evidence of large to moderate (OR>1.2) causal associations of vitamin D on a very wide range of health outcomes. These included SBP, DBP, hypertension, T2D, IHD, BMI, depression, non-vertebral fracture and all-cause mortality which have previously been proposed to be influenced by low vitamin D levels. Further, even larger studies, probably involving the joint analysis of data from several large biobanks with future IVs that explain a higher proportion of the trait variance, will be required to exclude smaller causal effects which could have public health importance because of the high population prevalence of low vitamin D levels in some populations.

TABLE OF CONTENTS

List of Tables

List of figures

Chapter I: Introduction – Vitamin D

1.1 Vitamin D

1.1.1 Vitamin D and its source/synthesis

Vitamin D is a group of fat-soluble secosteroids which help in enhancing intestinal absorption of calcium, iron, magnesium, phosphate and zinc. The most common forms of vitamin D are vitamin D2 (ergocalciferol) and vitamin D3 (cholecalciferol). The natural form of vitamin D in all animals, including human, is vitamin D3, while vitamin D2 is a synthetic product derived by irradiation of plant sterols/ergosterols (1).

A few food types are good sources of vitamin D, including oily fish (e.g. salmon, sardines and mackerel), eggs, fortified fat spreads, fortified breakfast cereals and some powdered milks (2). Another food source is vitamin D supplements. Apart from dietary sources and supplements, the majority of vitamin D we need is synthesized in our skin (**Table 1**).

Table 1. Dietary and supplemental sources of Vitamin D.

Source	Vitamin D content
Natural sources	
Salmon	
Fresh, wild (3.5 oz)	600-1000 IU of Vitamin D3
Fresh, farmed (3.5 oz)	100-250 IU of vitamin D3 or D2
Canned (3.5 oz)	300-600 IU of vitamin D3
Sardines, canned (3.5 oz)	300 IU of vitamin D3
Mackerel, canned (3.5 oz)	250 IU of vitamin D3
Tuna, canned (3.6 oz)	230 IU of vitamin D3
Cod liver oil (1 tsp)	400-1000 IU of vitamin D3
Shiitake mushrooms	

Fresh (3.5 oz)	100 IU of vitamin D3
Sun-dried (3.5 oz)	1600 IU of vitamin D3
Egg yolk	20 IU of Vitamin D3 or D2
Exposure to sunlight, ultraviolet B radiation (0.5 minimal erythema dose)	3000 IU of vitamin D3
Fortified foods	
Fortified milk	100 IU/8 oz, usually vitamin D3
Fortified orange juice	100 IU/8 oz vitamin D3
Infant formulas	100 IU/8 oz vitamin D3
Fortified yogurts	100 IU/8 oz, usually vitamin D3
Fortified butter	50 IU/3.5 oz, usually vitamin D3
Fortified cheese	100 IU/3 oz, usually vitamin D3
Fortified breakfast cereals	100 IU/serving, usually vitamin D3
Supplements	
Vitamin D2	50,000 IU/capsule
Drisdol (vitamin D2) liquid supplements	8000 IU/ml
Multivitamin	400 IU vitamin D, D2, or D3
Vitamin D3	400, 800, 1000, and 2000 IU

IU, international unit (25 ng); oz, ounces (28.3 grams).

Source: reproduced with permission from reference 267, Copyright Massachusetts Medical Society.

Oxidation of cholesterol to 7-dehydrocholesterol (7-DHC) is the initial step in vitamin D synthesis. 7-DHC is then stored principally in the cell membranes of keratinocytes and fibroblasts in the epidermis of skin, and is incorporated within the fatty acid hydrocarbon side chain and polar head group of the triglycerides (3). When exposed to sunlight, ultraviolet B radiation (UVB) (290-315 nm) is absorbed by epidermal cells, causing an activation of double bonds of 7-DHC (rearrange and open up the B ring) to form pre-vitamin D3 through a photolysis procedure. Once formed, pre-vitamin D3 rapidly converts to vitamin D3 via a 1-7 antarafacial sigma shift of a hydrogen from C-19 to C-9, which causes rearrangement of double bonds to form vitamin D3 (Figure 1) (4). All of the above steps of vitamin D synthesis happen in living cells of skin. Vitamin D3 is then ejected out of the plasma membrane into the extracellular space, binds with vitamin D binding protein (DBP) for transport to the liver (3).

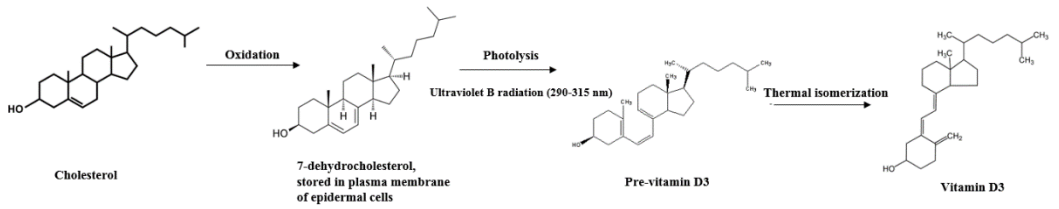


Figure 1. Synthesis of vitamin D.

Several factors could hinder vitamin D synthesis. Any physical factors that attenuate UVB exposure significantly reduce vitamin D3 synthesis in skin (e.g., clothing, sunscreens, glass shielding) (5). In addition, certain biological factors also inhibit cutaneous vitamin D synthesis and bioavailability. Increased skin pigmentation cuts the number of photons that reach the lower living cellular layer, where vitamin D3 is synthesized, and thus impact its availability. It is reported that under same degree of UVB radiation, significant ethnic group serum 25-hydroxyvitamin D (25(OH)D) level differences were observed between participants of multiple ethnicity with similar baseline 25(OH)D concentration (6). In recent study, obese subjects were found to have significantly lower basal 25(OH)D levels and higher parathyroid hormone concentrations compared with matched controls, which was caused by vitamin D's deposition in body fat compartments (7). Fat malabsorption syndromes, such as Crohn's disease, could possibility cause vitamin D insufficiency and deficiency (8). Age is another risk factor. With age increasing, the ability of the skin to synthesize vitamin D3 decreases and a shortage of intake vitamin D from food also exists. Compared with young adults, a 70-year-old produces approximately 4 times less vitamin D via cutaneous synthesis (9).

1.1.2 Metabolism and Physiologic Functions

To exert its biological function, vitamin D needs to be converted to its relevant active form first. Binding with DBP, vitamin D is transported to the liver in the serum. Then with cytochrome P450 enzyme 25-hydroxylase (CYP2R1), vitamin D is converted into 25(OH)D with a hydroxyl group adding on carbon 25 (10). Consequently, bound

to DBP, 25(OD)D is transferred to kidney tubules and with the aid of mitochondrial cytochrome P450 enzyme 25-hydroxyvitamin D-1α-hydroxylase (CYP27B1), 25(OH)D is converted to 1α,25-dihydroxyvitamin D (1,25$(OH)_2$D, or calcitriol) with a hydroxylation at position 1α (11). 1,25$(OH)_2$D is the active metabolite which is responsible for most biological functions of vitamin D (Figure 2).

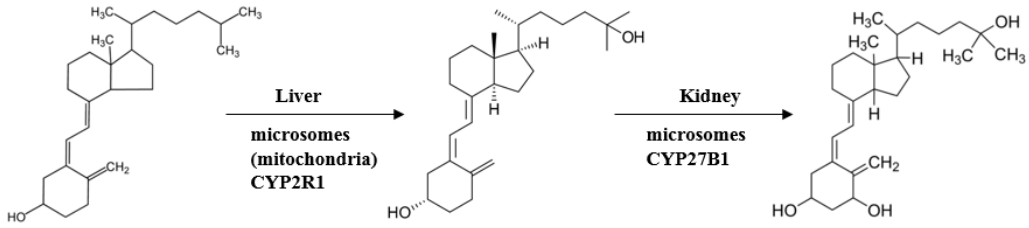


Figure 2. Activation of vitamin D

Source: reproduced from reference 13, with permission from Oxford University press (license number 4377770293884).

The main biological function of vitamin D involves calcium and phosphorus homeostasis. It was found that without vitamin D, only 10% to 15% of dietary calcium and about 60% of phosphorus is absorbed, while with vitamin D, calcium and phosphorus intestinal absorption are increased to around 40% and 80% respectively (12). Three metabolic pathways mediated by vitamin D are responsible for maintenance of serum calcium concentration. First, vitamin D impacts the intestinal absorption of calcium. Second, in absence of diet calcium, vitamin D stimulates transfer of calcium from bone to serum by osteoclastogenesis and activation of resting osteoclasts together with parathyroid hormone (13). Third, interacting with parathyroid hormone, the two hormones together can stimulate the reabsorption of the last 1% of filtered load of calcium in distal the renal tubule (14).

Apart from its role in calcium metabolism, vitamin D has also been identified to be involved in regulation of gene transcription, through its binding to the Vitamin D receptor (VDR). When vitamin D binds to VDR, it forms a heterodimer with the retinoid X receptor (RXR). This heterodimer binds to the vitamin D response element

(DRE) in the promoter of vitamin D response genes and interacts with other transcription factors (including co-activator proteins, such as SRC-1). As a consequence, DNA structure is remodeled through bend and phosphorylation on serine-205 occurs, and thus gene transcription could either be upregulated or downregulated (Figure 3) (13, 14). Expression of the *VDR* gene is not only found in calcium homeostasis related cells, such as enterocytes and distal renal tubule cells, but also in pro-myelocytes, lymphocytes, colon cells, pituitary gland cells and ovarian cells (13), which may implicate a regulatory function of vitamin D in many tissues and explain the impact of vitamin D on multiple health outcomes. Genes involved in calcium endocrinology were the first to be identified to have DRE regions, such as *Parathyroid Hormone Gene* (*PTH*) (15). However, DREs are also identified in a number of genes not linked to calcium metabolism. Cao X. et al. cloned the promoter of the avian beta 3 integrin gene, and identified the classic DRE sequence 756 to 770 bases upstream of beta 3 transcriptional start point (16). DRE was also identified in murine, rat and human fibronectin gene promoters (17).

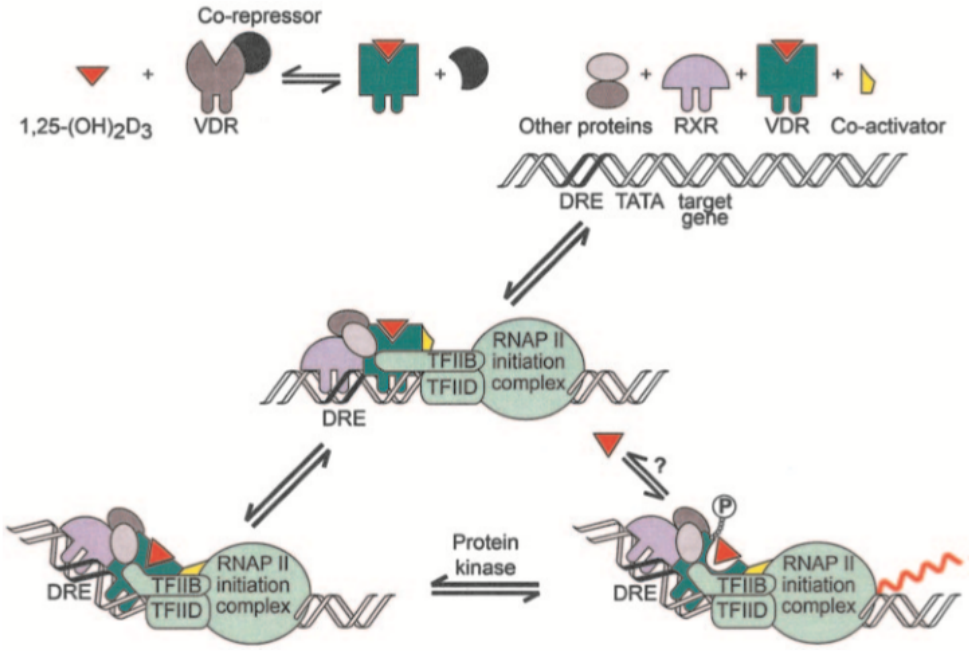


Figure 3. Regulation of gene expression by vitamin D.

$1,25(OH)_2D$, acting through its receptor, VDR. The result of regulation may be either suppression or activation. RXR, retinoid X receptor; DRE, vitamin D responsive element; TFIIB, transcription factor

IIB; TFIID, transcription factor IID; RNAP, RNA polymerase.

Source: reproduced from reference 13, with permission from Oxford University Press (reference number: 4377770613631).

1.1.3 Vitamin D deficiency

There is no consensus on optimal serum 25(OH)D concentrations, which is conventionally used to measure vitamin D status. Typically, vitamin D deficiency is considered when 25(OH)D concentration is lower than 20 ng/mL (50 nmol/L). Considering PTH levels and calcium absorption, a 25(OH) D level above 30 ng/mL (75 nmol/L) is considered to be sufficient (5), whereas a level between 20 ng/mL and 30 ng/mL is considered as insufficient. Through a systematic review of literature, Bischoff-Ferrari HA et al. suggested that for a wide variety of health outcomes (e.g. bone mineral density, dental health, colorectal cancer), the advantageous 25(OH)D levels begin at 30 ng/mL, and a higher level between 36-40 ng/mL would control disease risk further (18).

Although vitamin D can be synthesized by sun exposure, vitamin D deficiency is still highly prevalent all around the world, partly due to the factors which can hinder vitamin D synthesis, as have been detailed above. Overall, around 1 billion people worldwide were estimated to be vitamin D deficient or insufficient (12). For a view of vitamin D status around the world among adults, please see Figure 4 (19). In the US, the prevalence of 25(OH)D of less than 10 ng/mL was 6%, and only 67.2 % adults had 25(OH)D of 20 ng/mL or more. In the same study the proportion of adults with 25(OH)D concentration > 30ng/mL was only 23% (20, 21). Meanwhile, vitamin D deficiency is highly prevalent in Europe as well. In a study in French general adult population, 14% of participants were found to have a 25(OH)D concentration of lower than 12 ng/mL (22). In northwest Russia, the prevalence of vitamin D insufficiency and deficiency was estimated to be 37.5% and 45.7% respectively (23). Furthermore, in a British birth cohort of 7434 whites at 45 years old, the prevalence of 25(OH)D lower than 10 ng/mL (25 nmol/L), 16 ng/mL (40 nmol/L) and 30 ng/mL (75 nmol/L) was found to be 15.5%, 46.6% and 87.1% in winter and spring; while, the proportions were 3.2%, 15.4% and 60.9% during summer and fall (24).

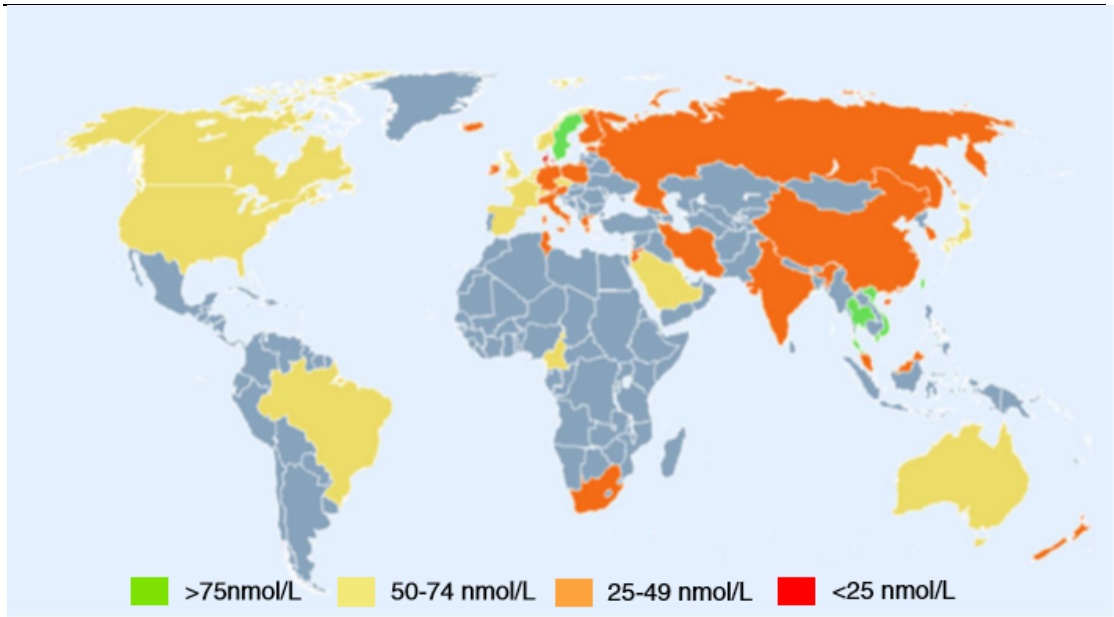


Figure 4. Vitamin D status in adults (> 18 years) around the world when available.

Winter values were used to calculate the mean 25(OH)D levels.

Source: reproduced from reference 19, with permission from Springer Nature (license number: 4377770985262).

1.2 Vitamin D in relation to health outcomes

1.2.1 Overview of association between vitamin D and health outcomes

Based on findings of previous systematic reviews and meta-analyses, my principal supervisor, Dr. Evropi Theodoratou, and colleagues conducted an umbrella review, which had been completed and published before my entry into the PhD project (25). Through 11 October 2013, they included 107 systematic literature reviews, 74 meta-analyses of observational studies of plasma vitamin D concentrations and 87 meta-analyses of RCTs of vitamin D supplementation. They found that associations between vitamin D and a total of 137 health outcomes had been explored by previous systematic reviews and meta-analyses which were published in the literature. They concluded that an association between vitamin D concentrations and birth weight, dental caries in children, maternal vitamin D concentrations at term, and parathyroid hormone concentrations in patients with chronic kidney disease requiring dialysis were probable. Furthermore, relationships between vitamin D and 22 health-related outcomes were

suggestive, including colorectal cancer, cardiovascular disease, stroke, cognition, type 2 diabetes and metabolic syndrome prevalence.

In order to form a comprehensive idea of which health outcomes have been studied in association with vitamin D, I firstly updated the aforementioned umbrella review by Dr. Evropi Theodoratou, using similar literature search terms and inclusion/exclusion criteria. For the new outcomes, which were not covered by the old umbrella review, I will briefly describe the results from the newly included studies in section **1.2.2**. For the outcomes that have been covered by the old umbrella review, I will describe evidence for several high prevalence outcomes based on the studies that were included in the old review in section **1.2.3**.

1.2.2 Association between vitamin D and health outcomes – an updated review

1.2.2.1 Search strategy, inclusion/exclusion criteria

I searched Medline and Embase for any systematic reviews or meta-analyses which studied the association between vitamin D and health outcomes and were published between November 2013 and March 2018, using the search algorithm listed in **Table 2**.

Two types of studies were included in this updated review: meta-analyses of observational associations between circulating vitamin D concentrations and any clinical outcome; meta-analyses of randomized controlled trials (RCTs) assessing the effect of vitamin D supplements. Only studies on humanity were eligible for inclusion. Letter, correspondence, conference abstracts, and articles not written in English were excluded. In addition, the following type of studies were also excluded from the review: studies on vitamin D related genetic variants; meta-analyses assessing dietary vitamin D intake or UVB exposure; studies which treat vitamin D level as outcome rather than exposure; meta-analysis of RCTs in which the treatment arm combined vitamin D with calcium or other vitamins or compounds versus placebo. However, if the treatment arum and control arm included the same additional compound (e.g. vitamin D plus calcium vs calcium), the study was included in my review.

Table 2. Search strategy in the update review for Medline and Embase

Medline	
Vitamin D terms	Vitamin d/ or 25-OHD.mp. or 25 hydroxyvitamin D.mp. or cholecalciferol/ or colecalciferol.mp. or hydroxycholecalciferols/ or hydroxycolecalciferols.mp. or calcifediol/ or dihydroxycholecalciferols/ or dihydroxycolecalciferols.mp. or calcitriol/ or 24,25-dihydroxyvitamin d 3/ or 24,25-OH2 D3.mp. or ergocalciferols/ or dihydrotachysterol/ or 25-hydroxyvitamin d 2/ or 25- OHD2.mp. or 1,25-dihydroxyvitamin d.mp. or 1,25-OH2 D.mp. or 1,25- dihydroxyvitamin d2.mp. or 1,25-dihydroxyergocalciferol.mp. or 1,25-OH2 D2.mp. or 1,25-dihydroxyvitamin d3.mp. or 1,25-OH2 D3.mp. or ergocalciferols/ or vitamin D2.mp. or vitamin D 2.mp. or vitamin D3.mp. or vitamin D 3.mp
	AND
Review terms	Review Literature as Topic/ or systematic review.mp or meta-analysis/ or metaanalysis.mp
Embase	
Vitamin D terms	Vitamin d/ or 25-OHD.mp. or 25 hydroxyvitamin D/ or colecalciferol/ or cholecalciferol.mp. or hydroxycolecalciferols/ or hydroxycholecalciferols.mp. or calcifediol/ or dihydroxycolecalciferols/ or dihydroxycholecalciferols.mp. or calcitriol/ or 24,25 dihydroxycolecalciferol/ or 24,25-OH2 D3.mp. or ergocalciferol/ or dihydrotachysterol/ or 25-hydroxyvitamin d 2.mp. or 25- OHD2.mp. or 1,25-dihydroxyvitamin d.mp. or 1,25-OH2 D.mp. or 1,25 dihydroxyergocalciferol/ or 1,25-dihydroxyvitamin d2.mp. or 1,25-OH2 D2.mp. or 1,25-dihydroxyvitamin d3.mp. or 1,25-OH2 D3.mp. or ergocalciferol/ or vitamin D2.mp. or vitamin D 2.mp. or vitamin D3.mp. or vitamin D 3.mp
	AND
Review terms	systematic review / or systematic review.mp or meta-analysis/ or metaanalysis.mp

Note: /, mapped afore term to subject heading; mp, including title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier and synonyms matching.

1.2.2.2 Literature search and review results

Medline and Embase were accessed on 21st March, 2018. Terms and algorithms from Table 2 were used in the literature search. 416 references were identified from Medline, and 1645 references were identified from Embase. After deletion of duplications, 1702 references were retained, which went to a subsequent review process. During title review, 1107 references were deleted. 208 were deleted in the abstract review, and thus we downloaded full text for 315 articles and did full text review for them. In the full text review, 141 studies were excluded because they were on outcomes which had been covered by the previous umbrella review, 4 studies were umbrella reviews of other systematic reviews and meta-analyses, 3 studies used vitamin D as outcomes, 6 studies explored vitamin D intake, 29 studies were just systematic reviews without any quantitative meta-analysis, 16 studies were meta-analyses of trials, however, the intervention strategy did not meet our inclusion criteria, and 21 studies were excluded due to other reasons. As a result, a total of 95 studies were included in the final dataset (**Figure 5**).

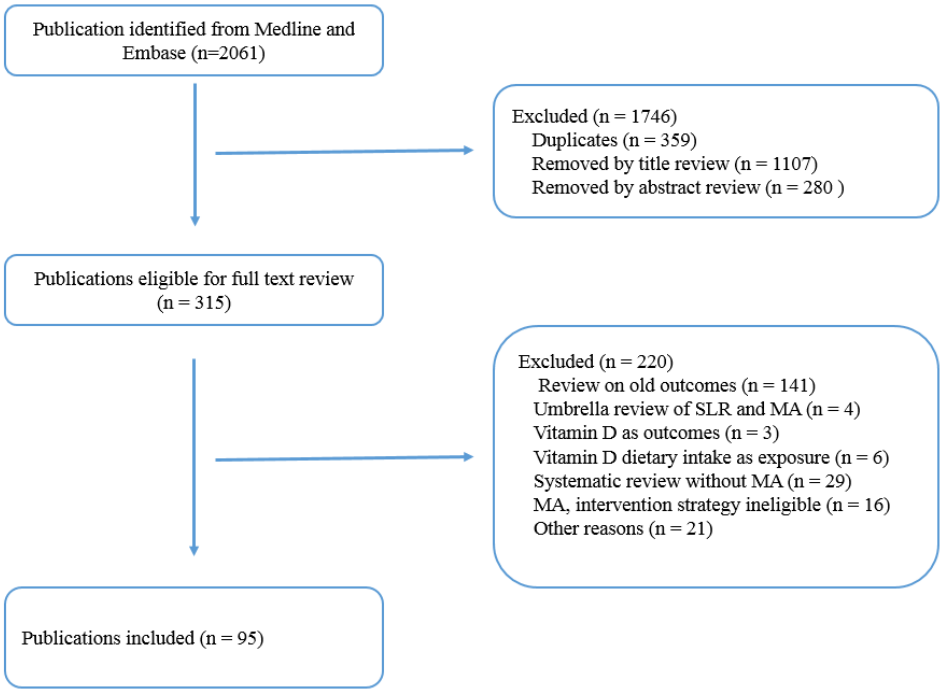


Figure 5. Flowchart of updated umbrella review for vitamin D.

SLR, systematic literature review; MA, meta-analysis.

1.2.2.3 Summary results of the updated systematic review

Among the included studies there were 66 meta-analyses of observational studies, and 29 meta-analyses of RCTs. For a summary of the results, please see **Table 3** and **Table 4**.

Cancer outcomes

The risk for basal cell melanoma, cutaneous melanoma, non-melanoma skin cancer and squamous cell cancer was explored and the associations for basal cell melanoma (RR = 1.82, 95% CI: 1.38-2.40) and non-melanoma skin cancer (RR = 1.64, 95% CI: 1.02-1.65) reached statistical significance level (26).

Table 3. General characteristics of meta-analyses of observational studies of vitamin D on health outcomes.

Outcome	Biomarker	Meta-analysis metric	Comparison	No of studies in each MA	No of events	total N	Relative Risk (95% CI)	P value	I^2 (95% CI)	*P* value for heterogeneity
Cancer outcomes										
basal cell melanoma (26)	25(OH)D	RR	highest quantile vs lowest quantile	5	NS	NS	1.82 (1.38, 2.40)	NS	0	NS
bladder cancer (27)	25(OH)D	RR	high vs low	5	2238	89610	0.75 (0.65, 0.87)	<0.001	0	0.42
bladder carcinoma (28)	25(OH)D	RR	low vs high	7	NS	NS	1.34 (1.17, 1.53)	<0.001	0	0.53
breast cancer, disease free survival (29)	25(OH)D	HR	high vs low	4	NS	NS	0.42 (0.29, 0.62)	<0.001	29	0.24
breast cancer, overall mortality (30)	25(OH)D	RR	high vs low	6	NS	NS	0.61 (0.48, 0.79)	NS	35.9	0.17
breast cancer, overall mortality (31)	25(OH)D	HR	high vs low	5	471	4443	0.56 (0.4, 0.7)	<0.001	NS	0.35
breast cancer, overall mortality (32)	25(OH)D	HR	high vs low	5	622	4413	0.62 (0.49, 0.78)	NS	4.6	0.2

breast cancer, overall survival (29)	25(OH)D	HR	high vs low	6	NS	NS	0.63 (0.51, 0.77)	<0.001	14	0.32
breast cancer, specific mortality (30)	25(OH)D	RR	high vs low	4	NS	NS	0.58 (0.40, 0.85)	NS	26.7	0.251
breast cancer, specific mortality (29)	25(OH)D	HR	high vs low	4	NS	NS	0.65 (0.44, 0.98)	0.04	45	0.14
breast cancer, specific mortality (32)	25(OH)D	HR	high vs low	3	194	2636	0.57 (0.38, 0.84)	NS	17	0.12
colorectal cancer, overall mortality (32)	25(OH)D	HR	high vs low	5	1214	2330	0.71 (0.55, 0.91)	NS	29	0.12
colorectal cancer, overall survival (29)	25(OH)D	HR	high vs low	5	NS	NS	0.55 (0.33, 0.91)	0.02	89	<0.001
colorectal cancer, specific mortality (29)	25(OH)D	HR	high vs low	3	NS	NS	0.65 (0.47, 0.88)	0.005	6	0.35
colorectal cancer, specific mortality (32)	25(OH)D	HR	high vs low	3	566	1558	0.65 (0.49, 0.86)	NS	0	0.34
colorectal cancer, specific mortality (33)	25(OH)D	HR	high vs low	4	809	2559	0.63 (0.5, 0.8)	<0.001	NS	0.5
cutaneous melanoma (26)	25(OH)D	RR	highest quantile vs lowest quantile	4	NS	NS	1.46 (0.60, 3.53)	NS	54	NS

gastric cancer (34)	25(OH)D	OR	high vs low	3	NS	NS	0.92 (0.74, 1.14)	0.42	22.5	0.27
haematological malignancies, disease free survival (35)	25(OH)D	HR	low vs high	5	NS	NS	1.45 (1.25, 1.70)	<0.001	0	0.558
haematological malignancies, overall survival (35)	25(OH)D	HR	low vs high	7	NS	NS	1.85 (1.54, 2.23)	<0.001	0	0.747
lung cancer, overall mortality (36)	25(OH)D	OR	high vs low	3	NS	NS	0.39 (0.28, 0.54)	<0.01	92.5	<0.001
lung cancer, overall mortality (37)	25(OH)D	RR	high vs low	4	NS	NS	0.76 (0.61, 0.94)	0.014	96.1	<0.001
lung cancer, overall survival (29)	25(OH)D	HR	high vs low	4	NS	NS	0.75 (0.30, 1.86)	0.54	93	<0.001
lung cancer, overall survival (36)	25(OH)D	OR	high vs low	4	NS	NS	1.01 (0.87, 1.18)	0.87	19.4	0.293
lung cancer, overall survival (37)	25(OH)D	RR	high vs low	5	NS	NS	1.01 (0.88, 1.16)	0.872	0	0.437
lymphoma, disease free survival (29)	25(OH)D	HR	high vs low	6	NS	NS	0.80 (0.65, 0.98)	0.04	0	0.61
lymphoma, overall survival (29)	25(OH)D	HR	high vs low	7	NS	NS	0.48 (0.36, 0.64)	<0.001	0	0.86
lymphoma, specific mortality (29)	25(OH)D	HR	high vs low	6	NS	NS	0.50 (0.36, 0.68)	<0.001	0	0.83

non-melanoma skin cancer (26)	25(OH)D	RR	highest quantile vs lowest quantile	2	NS	NS	1.64 (1.02, 2.65)	NS	81	NS
squamous cell cancer (26)	25(OH)D	RR	highest quantile vs lowest quantile	4	NS	NS	1.68 (0.44, 6.39)	NS	81	NS
thyroid cancer (38)	25(OH)D	OR	low vs high	9	NS	NS	1.42 (1.17, 1.73)	NS	35.6	0.134
thyroid cancer (38)	25(OH)D	SMD of VD	cases vs controls	7	NS	NS	-0.20 (-0.36, -0.03)	NS	55.4	0.037
Cardiovascular outcomes										
atrial fibrillation (39)	25(OH)D	OR	low vs high	5	NS	NS	1.31 (1.06, 1.62)	0.01	60	0.04
Autoimmune diseases										
Grave's disease (40)	25(OH)D	SMD of VD	cases vs controls	26	1748	3596	-0.77 (-1.12, -0.42)	NS	95.5	<0.001

Grave's disease (40)	25(OH)D	OR	low vs high	12	NS	NS	2.24 (1.31, 3.81)	NS	84.1	<0.001
inflammatory bowel disease (41)	25(OH)D	OR	cases vs controls	14	938	1891	1.64 (1.30, 2.08)	<0.001	7	0.37
systemic sclerosis (42)	25(OH)D	SMD of VD	cases vs controls	6	554	875	-8.72 (-10.11, -7.32)	<0.001	48	0.09
ulcerative colitis (41)	25(OH)D	OR	cases vs controls	7	177	539	2.28 (1.18, 4.41)	0.01	41	0.12
ulcerative colitis (43)	25(OH)D	MD of VD	cases vs controls	5	235	517	-6.72 (-16.49, 3.05)	0.1777	NS	NS
ulcerative colitis (44)	25(OH)D	SMD of VD	cases vs controls	8	296	810	-0.50 (-0.85, -0.15)	NS	77.5	<0.001
ulcerative colitis (44)	25(OH)D	OR	cases vs controls	3	NS	NS	2.02 (1.13, 3.60)	NS	0	0.773
Mental health disorders and Cognitive traits										
autism spectrum disorder (45)	25(OH)D	MD of VD	cases vs controls	11	NS	1652	-8.63 (-13.17, -4.09)	0.0002	98	<0.0001

cognitive impairment (46)	25(OH)D	OR	low vs high	NS	NS	NS	1.56 (1.05, 2.33)	0.028	NS	NS
cognitive impairment (46)	25(OH)D	MD	cases vs controls	7	1179	6068	-6.83 (-11.36, -2.30)	<0.001	99	NS
dementia (47)	25(OH)D	OR	low vs high	5	NS	NS	1.49 (1.09, 1.88)	NS	0	0.46
dementia (48)	25(OH)D	OR	low vs high	5	NS	18933	1.54 (1.19, 1.99)	NS	20	NS
schizophrenia (49)	25(OH)D	MD of VD	cases vs controls	14	NS	NS	-5.91 (-10.68, -1.14)	NS	97.6	<0.001
schizophrenia (49)	25(OH)D	OR	low vs high	8	NS	NS	2.16 (1.32, 3.56)	NS	46.8	0.068
Infectious diseases										
clostridium difficile infection (50)	25(OH)D	WMD of VD	cases vs controls	3	NS	NS	-3.54 (-6.89, -0.39)	NS	48	0.15
clostridium difficile infection recurrence (50)	25(OH)D	OR	low vs high	4	NS	NS	1.26 (0.56, 2.83)	NS	63	0.04
clostridium difficile infection severity (50)	25(OH)D	OR	low vs high	3	NS	NS	1.61 (1.02, 2.53)	NS	1	0.37

mortality among pneumonia patients (51)	25(OH)D	RR	low vs high	6	NS	NS	2.59 (1.32, 5.08)	0.005	71.4	<0.001
Metabolic disorders										
antiphospholipid syndrome (52)	25(OH)D	SMD of VD	cases vs controls	3	NS	NS	-3.605 (-5.449, -1.761)	<0.001	0 (0, 99.9)	NS
antiphospholipid syndrome (52)	25(OH)D	OR	cases vs controls	4	NS	NS	3.063 (2.120, 4.426)	<0.001	24.7 (0, 90.3)	NS
diabetic peripheral neuropathy (Asian diabetes patients) (53)	25(OH)D	OR	low vs high	4	NS	NS	1.22 (1.17, 1.27)	NS	0	0.77
diabetic nephropathy (diabetes patients) (54)	25(OH)D	OR	low vs high	6	NS	NS	1.80 (1.25, 2.59)	0.002	59.4	0.031
diabetic peripheral neuropathy (diabetes patients) (53)	25(OH)D	SMD of VD	cases vs controls	10		1368	-1.12 (-1.58, -0.65)	NS	94.1	<0.001
diabetic peripheral neuropathy (T2D patients) (55)	25(OH)D	WMD of VD	cases vs controls	5	NS	606	-6.36 (-8.57, -4.14)	<0.001	56	0.06
diabetic peripheral neuropathy (T2D patients) (55)	25(OH)D	OR	low vs high	4	NS	NS	2.88 (1.84, 4.50)	<0.001	0	0.47
diabetic retinopathy (T2D patients) (56)	25(OH)D	OR	low vs high	8	2348	13435	2.03 (1.07, 3.86)	0.03	96	<0.001

obesity (57)	25(OH)D	RR	obesity vs entrophic	21	NS	NS	1.35 (1.21, 1.50)	NS	87.3	<0.001
obesity (58)	25(OH)D	RR	cases vs controls	12	NS	NS	1.52 (1.33, 1.74)	NS	89.3	<0.001
obesity (59)	25(OH)D	OR	cases vs controls	15	3867	13209	3.43 (2.33, 5.06)	NS	81.2	<0.001
overweight (57)	25(OH)D	RR	obesity vs overweight	19	NS	NS	1.21 (1.14, 1.29)	NS	66.4	<0.001
type 2 diabetes (old adults) (60)	25(OH)D	RR	low vs high	6	1320	13563	1.31 (1.11, 1.54)	0.001	37	NS
Urolithiasis (61)	1,25(OH)D	MD	cases vs controls	23	2189	4237	10.19 (4.31, 16.07)	0.0007	97	<0.001
Urolithiasis (61)	25(OH)D	MD	cases vs controls	12	1934	20850	0.88 (-1.04, 2.80)	0.37	84	<0.001
Neonatal/infant/child related outcomes										
allergic sensitization in children (maternal VD) (62)	25(OH)D	OR	high vs low	9	NS	NS	1.00 (0.95, 1.06)	0.962	0	0.739
FEVA in children (maternal VD) (62)	25(OH)D	OR	high vs low	4	NS	NS	0.07 (-0.01, 0.15)	0.104	0	0.876

FVC in children (maternal VD) (62)	25(OH)D	OR	high vs low	4	NS	NS	0.05 (-0.03, 0.13)	0.192	NS	NS
lower respiratory tract infection (children) (63)	25(OH)D	OR	cases vs controls	5	197	550	3.29 (1.27, 8.56)	0.01	51	0.09
lower respiratory tract infection (children) (63)	25(OH)D	MD of VD	cases vs controls	6	NS	578	-8.75 (-15.70, -1.80)	0.01	66	0.01
respiratory tract infection in children (maternal VD) (62)	25(OH)D	OR	high vs low	13	NS	NS	0.64 (0.47, 0.87)	0.005	82.9	<0.001
Pregnancy related outcomes										
biochemical pregnancy (64)	25(OH)D	OR	high vs low	5	975	1700	1.34 (1.04, 1.73)	0.03	21	0.28
clinical pregnancy (64)	25(OH)D	OR	high vs low	11	1305	2700	1.46 (1.05, 2.02)	0.02	61	0.004
clinical pregnancy (in vitro fertilization) (65)	25(OH)D	RR	low vs high	4	631	1139	0.92 (0.73, 1.16)	NS	61	0.053
live birth (64)	25(OH)D	OR	high vs low	7	895	2026	1.33 (1.08, 1.65)	0.009	5	0.39
live birth rate (in vitro fertilization) (65)	25(OH)D	RR	low vs high	3	243	655	0.75 (0.61, 0.93)	NS	0	0.357
miscarriage (64)	25(OH)D	OR	high vs low	6	240	1635	1.12 (0.81, 1.54)	0.49	0	0.76

polycystic ovary syndrome (66)	25(OH)D	SMD of VD	cases vs controls	14	1150	2262	-0.64 (-1.12, -0.15)	0.01	96	<0.001
polycystic ovary syndrome (67)	25(OH)D	SMD of VD	cases vs controls	10	1130	1879	-0.45 (-1.68, 0.79)	0.48	99	<0.001
polycystic ovary syndrome (68)	25(OH)D	SMD of VD	cases vs controls	10	NS	NS	-0.86 (-1.46, -0.26)	0.005	97.2	<0.001
spontaneous abortion (69)	25(OH)D	RR	low vs high	3	NS	NS	1.04 (0.95, 1.13)	NS	0	0.71
spontaneous preterm birth (<35-37 weeks) (69)	25(OH)D	RR	low vs high	4	NS	NS	1.11 (0.75, 1.65)	NS	38	0.184
still birth (69)	25(OH)D	RR	low vs high	2	NS	NS	1.02 (0.96, 1.09)	NS	0	0.65
Musculoskeletal and related outcomes										
chronic widespread pain (70)	25(OH)D	OR	cases vs controls	9	NS	NS	1.63 (1.20, 2.23)	NS	37.8	0.117
fibromyalgia (71)	25(OH)D	SMD of VD	cases vs controls	12	851	1713	-0.56 (-1.05, -0.08)	NS	95.5	<0.001
low back pain (72)	25(OH)D	OR	low vs high	19	NS	NS	1.60 (1.20, 2.12)	0.001	84.9	NS

stress fracture (military population) (73)	25(OH)D	MD	cases vs controls	8	761	2634	-2.44 (-4.05, -0.84)	0.003	53	0.04
walking speed in elderly (74)	25(OH)D	MD	VDD vs NVD	7	3313	7623	-0.08 (-0.09, -0.07)	<0.001	85	NS
Other outcomes										
30-day mortality (patients in ICU) (75)	25(OH)D	RR	low vs high	7	2857	2857	1.42 (1.00, 2.02)	0.05	29	0.21
advanced liver fibrosis (76)	25(OH)D	OR	low vs high (10 ng/mL)	2	NS	NS	2.37 (1.20, 4.72)	NS	NS	NS
advanced liver fibrosis (76)	25(OH)D	OR	low vs high (20 ng/mL)	3	NS	NS	1.44 (0.99, 2.12)	NS	NS	NS
advanced liver fibrosis (76)	25(OH)D	OR	low vs high (30 ng/mL	3	NS	NS	2.22 (1.24, 3.97)	NS	NS	NS
aeroallergen specific Ig-E sensitization (77)	25(OH)D	OR	high vs low	4	NS	NS	0.889 (0.754, 1.048)	0.16	66.2	<0.001
aeroallergen specific Ig-E sensitization (adults) (77)	25(OH)D	OR	high vs low	3	NS	NS	1.263 (1.121, 1.423)	<0.001	0	0.89

aeroallergen specific Ig-E sensitization (children) (77)	25(OH)D	OR	high vs low	2	NS	NS	0.542 (0.433, 0.678)	<0.001	49.7	0.002
age-related macular degeneration (78)	25(OH)D	SMD of VD	cases vs controls	3	1126	9332	-0.15 (-0.41, 0.11)	0.272	NS	NS
age-related macular degeneration (78)	25(OH)D	OR	highest quintile vs lowest quintile	4	2784	26572	0.83 (0.71, 0.97)	0.019	78.4	NS
age-related macular degeneration (79)	25(OH)D	OR	low vs high	8	NS	NS	0.91 (0.69, 1.22)	0.12	79.7	<0.01
chronic obstructive pulmonary disease (80)	25(OH)D	SMD of VD	controls vs cases	9	NS	NS	0.60 (0.31, 0.89)	<0.001	88.5	<0.001
chronic obstructive pulmonary disease (81)	25(OH)D	SMD of VD	cases vs controls	13	1981	3264	-0.69 (-1.00, -0.38)	<0.001	94	<0.001
chronic obstructive pulmonary disease (81)	25(OH)D	OR	cases vs controls	12	3224	9923	1.77 (1.18, 2.64)	0.006	83	<0.001
fibrosis score (NAFLD) (82)	25(OH)D	MD of VD	high vs low score	6	NS	NS	0.88 (-2.65, 4.42)	0.65	64	0.62
frailty (83)	25(OH)D	OR	low vs high	7	NS	17815	1.27 (1.17, 1.38)	<0.001	59	0.02

infection (patients in ICU) (75)	25(OH)D	RR	low vs high	5	334	1965	1.49 (1.12, 1.99)	0.007	52	0.08
in-hospital mortality (patients in ICU) (75)	25(OH)D	RR	low vs high	4	2572	2572	1.76 (1.37, 2.26)	<0.001	0	0.83
mortality (critically ill patients) (84)	25(OH)D	OR	low vs high	6	NS	NS	1.76 (1.38, 2.23)	NS	2.3	0.402
NAFLD activity score (82)	25(OH)D	MD of VD	high vs low score	5	NS	NS	-0.93 (-2.45, 0.58)	0.23	0	0.64
non-alcoholic fatty liver disease (85)	25(OH)D	SMD of VD	cases vs controls	21	NS	NS	-0.76 (-0.97, -0.54)	NS	95.6	<0.001
non-alcoholic fatty liver disease (85)	25(OH)D	OR	low vs high	6	NS	NS	1.26 (1.15, 1.38)	NS	39.7	0.141
non-alcoholic steatohepatitis (85)	25(OH)D	SMD of VD	cases vs controls	4	NS	374	-1.30 (-2.37, -0.23)	0.02	94	<0.001
non-cardiovascular, non-cancer death (86)	25(OH)D	RR	bottom vs top thirds	10	2565	51561	1.34 (1.13, 1.60)	NS	49.3	0.038
obstructive sleep apnea (87)	25(OH)D	SMD of VD	controls vs mild OSA	5	NS	2381	-0.08 (-0.29, 0.13)	0.46	44	0.13
obstructive sleep apnea (87)	25(OH)D	SMD of VD	controls vs moderate OSA	6	NS	1922	-0.29 (-0.55, -0.04)	0.02	61	0.02

obstructive sleep apnea (87)	25(OH)D	SMD of VD	controls vs severe OSA	7	NS	1942	-0.56 (-0.85, -0.27)	0.0002	77	0.0002
Parkinson's disease (88)	25(OH)D	SMD of VD	cases vs controls	7	NS	5544	-16.88 (-33.54, -0.23)	0.05	99	<0.001
Parkinson's disease (88)	25(OH)D	OR	low vs high	2	385	870	1.50 (1.14, 1.97)	0.003	49	0.16
Parkinson's disease (89)	25(OH)D	MD of VD	cases vs controls	6	NS	1893	-11.55 (-12.23, -10.86)	<0.001	99	<0.001
Parkinson's disease (90)	25(OH)D	OR	low vs high	3	NS	NS	1.50 (1.31, 1.71)	NS	55.9	0.045
sepsis (hospitalized patients) (91)	25(OH)D	OR	low vs high	7	1452	7194	1.78 (1.55, 2.03)	<0.001	0	0.97
sepsis (patients in ICU) (75)	25(OH)D	RR	low vs high	7	854	3844	1.46 (1.27, 1.68)	<0.001	0	0.47

Table 4. General characteristics of meta-analyses of randomized controlled trials of vitamin D supplementation on health outcomes.

Outcome	Population	Type of metric (summary effect)	Units	MA model	No of studies in each MA	Reported summary effect (95% CI)	P value
Skeletal outcomes							
chromic nonspecific musculoskeletal pain (92)	general	SMD	NA	random	3	0.05 (-0.37, 0.46)	0.83
hand grip strength (93)	older adults	MD	kg	fixed	3	0.40 (-1.11, 1.92)	0.6
joint space width (94)	knee OA patients	SMD	NA	fixed	2	0.07 (-0.08, 0.23)	0.36
knee osteoarthritis, WMAC function (95)	knee OA patients	WMD	NA	NA	4	-1.87 (-2.58, -1.17)	NA
knee osteoarthritis, WMAC pain (94)	knee OA patients	SMD	NA	random	4	-0.32 (-0.63, -0.02)	0.04
knee osteoarthritis, WMAC pain (95)	knee OA patients	WMD	NA	NA	4	-1.65 (-2.16, -1.14)	NA
knee osteoarthritis, WMAC stiffness (95)	knee OA patients	WMD	NA	NA	3	0.03 (-0.17, 0.24)	NA
leg strength (93)	older adults	SMD	NA	fixed	7	0.09 (-0.05, 0.24)	0.22
pain score (96)	general	MD	NA	random	8	-0.57 (-1.00, -0.15)	0.007
physical performance (93)	older adults	SMD	NA	fixed	4	0.12 (-0.07, 0.30)	0.22
tibial cartilage volume (94)	knee OA patients	SMD	NA	fixed	2	0.12 (-0.05, 0.29)	0.15
time up and go (93)[a]	older adults	MD	seconds	fixed	7	-0.75 (-1.44, -0.07)	0.03
visual analog scale of pain intensity (97)	chronic widespread pain patients	MD	NA	random	4	0.46 (0.09, 0.89)	0.02
walking capacity (93)	older adults	SMD	NA	fixed	3	0.04 (-0.17, 0.24)	0.73
cardiovascular outcomes							

6-minute walk distance (98)	chronic heart failure patients	MD	meters	fixed	2	8.90 (-48.47, 66.26)	0.761
augmentation index (99)	general	SMD	NA	fixed	8	-0.15 (-0.32, 0.02)	0.08
augmentation index (100)	general	MD	NA	random	4	0.25 (-4.43, 4.92)	0.92
cardiovascular mortality (101)	CKD patients	RR	NA	random	6	0.79 (0.26, 2.38)	NA
cardiovascular serious adverse events (101)	CKD patients	RR	NA	random	8	1.20 (0.48, 2.99)	NA
diastolic blood pressure (102)	obese individuals	SMD	mmHg	random	5	0.124 (-0.003, 0.251)	0.055
endothelial function (103)	general	SMD	NA	random	14	0.08 (-0.06, 0.22)	0.28
fasting flow-mediated vasodilation (104)	general	WMD	% change	fixed	9	0.15 (-0.21, 0.51)	0.41
flow mediated dilation (105)	general	SMD	% change	random	11	1.27 (0.20, 2.34)	NA
flow mediated dilation (106)	general	MD	% change	random	8	0.96 (-0.14, 2.06)	0.09
incident CVD (107)	predialysis CKD patients	RR	NA	random	5	0.27 (0.13, 0.59)	0.001
interleukin-10 concentration (98)	chronic heart failure patients	MD	pg/mL	random	2	0.94 (-0.72, 2.59)	0.269
left ventricular ejection fraction (98)	chronic heart failure patients	MD	% change	random	4	4.11 (-0.91, 9.12)	0.109
N-terminal pro-B-type natriuretic peptide (98)	chronic heart failure patients	MD	pg/mL	random	3	-80.8 (-305.3, 143,7)	0.48
pulse wave velocity (99)	general	SMD	NA	fixed	10	-0.10 (-0.24, 0.04)	0.17
pulse wave velocity (100)	general	MD	NA	random	6	0.18 (-0.17, 0.52)	0.31
systolic blood pressure (102)	obese individuals	SMD	mmHg	random	5	0.239 (0.086, 0.391)	0.002

TNF-a concentration (98)	chronic heart failure patients	MD	pg/mL	random	3	-2.42 (-4.26, -0.57)	0.01
metabolic disorders							
adiponectin concentration (108)	general	MD	% change	random	7	4.45 (-3.04, 11.93)	0.244
body mass index (102)	obese individuals	SMD	kg/m2	random	4	0.097 (-0.113, 0.307)	0.365
fasting plasma glucose (109)	T2D patients	SMD	mg/dL	random	16	-6.7 (-11.0, -2.2)	0.003
fat mass (110)	general	WMD	kg	random	10	-0.03 (-0.63, 0.57)	0.92
fat mass (111)	general	SMD	kg	random	10	-0.014 (-0.355, 0.308)	0.934
fat mass (%) (111)	general	SMD	% change	random	12	0.051 (-0.098, 0.200)	0.503
HbA1c (109)	T2D patients	SMD	% change	random	18	-0.25 (-0.41, -0.09)	0.003
HDL-C (102)	obese individuals	SMD	mmol/L	random	4	0.110 (-0.056, 0.277)	0.194
HOMA-IR (102)	obese individuals	SMD	mL/min/kg	random	4	-0.078 (-0.221, 0.066)	0.288
HOMA-IR (109)	T2D patients	SMD	NA	random	8	-0.62 (-1.2, -0.05)	0.03
LDL-C (102)	obese individuals	SMD	mmol/L	random	4	0.338 (0.071, 0.605)	0.013
leptin concentration (108)	general	MD	% change	random	6	-4.51 (-25.13, 16.11)	0.668
triglycerides (102)	obese individuals	SMD	mmol/L	random	4	-0.236 (-0.497, 0.025)	0.076
weight (110)	general	WMD	kg	random	16	0.01 (-0.21, 0.23)	0.9
weight (102)	obese individuals	SMD	kg	random	5	0.097 (-0.113, 0.307)	0.365

Neonatal/infant/child related outcomes							
asthma exacerbation in children (112)	children	RR	NA	random	3	0.41 (0.27, 0.63)	<0.001
asthma exacerbations requiring treatment with systemic corticosteroids in children (113)	individuals with athma	RR	NA	random	7	0.74 (0.56, 0.97)	0.03
asthma exacerbation in children (114)	children	RR	NA	random	3	0.41 (0.27, 0.63)	<0.001
Other outcomes							
all-cause mortality (101)	CKD patients	RR	NA	random	13	0.84 (0.46, 1.52)	NA
C-reactive protein concentration (115)	overweight and obese individuals	SMD	mg/L	random	12	-0.11 (-0.27, 0.04)	0.15
C-reactive protein, Systemic inflammatory (116)	general	WMD	NA	random	8	-0.32 (-1.01, 0.36)	0.352
high-density CRP (117)	general	WMD	mg/L	random	10	-1.08 (-2.13, -0.03)	NA
hospital length of stay (118)	ICU patients	WMD	days	random	6	-3.11 (-10.04, 3.82)	0.38
ICU length of stay (118)	ICU patients	WMD	days	random	6	-1.42 (-3.78, 0.94)	0.24
interleukin 6 concentration (115)	overweight and obese individuals	SMD	pg/mL	random	5	0.10 (-0.43, 0.63)	0.71
interleukin 6, Systemic inflammatory (116)	general	WMD	NA	random	3	0.10 (-0.17, 0.37)	0.462
length of hospital stay (119)	ICU patients	SMD	NA	fixed	5	-0.06 (-0.22, 0.10)	0.47
length of ICU stay (119)	ICU patients	SMD	NA	fixed	5	-0.13 (-0.29, 0.03)	0.11
length of mechanical ventilation (119)	ICU patients	SMD	NA	fixed	3	-0.04 (-0.21, 0.13)	0.65
mechanical ventilator days (118)	ICU patients	WMD	days	random	4	-1.20 (-3.72, 1.33)	0.35
mortality (118)	ICU patients	RR	NA	random	6	0.84 (0.66, 1.06)	0.14
mortality (119)	ICU patients	OR	NA	fixed	5	0.70 (0.50, 0.98)	0.04

number of dominant follicles (120)	PCOS patients	OR	NA	fixed	4	2.34 (1.39, 3.92)	0.001
number of regular menstrual cycles (120)	PCOS patients	OR	NA	fixed	3	1.51 (0.80, 2.83)	0.2
PTH level (120)	PCOS patients	MD	NA	random	5	-13.23 (-23.30, -3.17)	0.01
TNF-a concentration (115)	overweight and obese individuals	SMD	pg/mL	random	8	-0.13 (-0.38, 0.12)	0.31
urine albumin to creatinine ratio (54)	T2D patients	WMD	NA	random	4	17.99 (-35.36, 71.33)	0.51

a. time up and go, a measurement for physical performance, which is the time that a person takes to rise from a chair, walk three meters, turn around, walk back to the chair.

The risk for bladder cancer was newly reported for 2 systematic reviews for observational studies. The risk for bladder cancer was reported to be 0.75 (95% CI: 0.65-0.87, P<0.001, high vs low level of vitamin D) (27), while another review on bladder carcinoma reported the risk to be 1.34 (95%CI: 1.17-1.53, P<0.001, low vs high level of vitamin D) (28).

Four studies included breast cancer mortality as their outcome, and the results were all statistically significant and consistent in direction: 0.58 (95% CI: 0.40-0.85), 0.56 (95% CI: 0.40-0.70), 0.57 (95% CI: 0.38-0.84), 0.65 (95% CI: 0.44-0.98), respectively (29-32). One study further analysed breast cancer overall survival (HR = 0.63, 95% CI: 0.51-0.77), and breast cancer disease-free survival (HR = 0.42, 95% CI: 0.29-0.62), all of which were statistically significant.

Colorectal cancer risk was studied in two articles, both of which were statistically significant (HR = 0.65, 95% CI: 0.47-0.88; HR = 0.65, 95% CI: 0.49-0.86) (29, 32). In addition, colorectal cancer overall survival was also found to be statistically significantly associated with vitamin D (HR = 0.55, 95% CI: 0.33-0.91) (29).

Only one study used gastric cancer as their outcome of interested, and the result of MA was not statistically significant (RR = 1.46, 95% CI: 0.60-3.53).

Lung cancer mortality was studies in two studies and the results were statistically significant and consistent (OR = 0.39, 95% CI: 0.28-0.54; RR = 0.76, 95% CI: 0.61-0.94) (36, 37). However, the results for lung cancer survival were not statistically significant in all three studies (29, 36, 37).

Overall survival and disease free survival of haematological malignancies were also found to be statistically significantly associated with vitamin D (HR = 1.85, 95% CI: 1.54-2.23 for overall survival; HR = 1.45, 95% CI: 1.25-1.70 for relapse-free survival) (35).

The evidence for risk of thyroid cancer was synthesized in one study. And it was found to be statistically significantly associated with vitamin D (OR = 1.42, 95% CI: 1.17-1.73) (38).

Cardiovascular outcomes

There was only one study of observational studies summarizing the evidence for cardiovascular outcomes. It studied the risk of atrial fibrillation and found it to be statistically significant (OR = 1.31, 95% CI: 1.06-1.62) (39).

Jiang et al. studied the effect of vitamin D supplementation in a MA of RCTs. They found that vitamin D intervention did not change 6-minute walk distance, left ventricular ejection fraction, N-terminal pro-B-type natriuretic peptide or interleukin-10 concentration profile. But it showed a statistically significant association with tumour necrosis factor-a level (MD = -2.42 pg/mL, 95% CI: -4.26~ -0.57) (98).

Arterial stiffness was studied by two MA of RCTs, presenting by augmentation index and pulse wave velocity. But neither were statistically significant (99, 100).

The association of vitamin D supplementation with blood pressure of obese individuals was studied, and systolic blood pressure was statistically significantly affected by the intervention (SMD = 0.239 mmHg, 95% CI: 0.086, 0.391) (102).

The association of vitamin D supplements with flow mediated dilation, a marker to access risk of CVD events, was explored in 3 MA studies of RCTs (104-106). The results were inconsistent between studies, and only one study found a statistically significant association (SMD = 1.27, 95% CI: 0.20-2.34) (105).

Among predialysis chronic kidney disease (CKD) patients, vitamin D supplements statistically significantly decreased incident cardiovascular disease (CVD) (RR = 0.27, 95% CI: 0.13-0.59) as reported by a MA of RCTs (107). Another MA of RCTs studied cardiovascular serious adverse events and CVD mortality in CKD patients, however the results were not statistically significant (101).

Autoimmune diseases

The associations between vitamin D concentration and ulcerative colitis were reported by three studies (41, 43, 44). Two of them reported statistically significant results (OR = 2.28, 95% CI: 1.18-4.41; OR= 2.02, 95% CI: 1.13-3.60) (41, 44).

By a MA of observational studies, risk Grave's disease was statistically significantly associated with serum vitamin D level (OR = 2.24, 95% CI: 1.31-3.81; SMD of vitamin D (cases vs controls): -0.77, 95% CI: -1.12 ~ -0.42) (40).

The risk of inflammatory bowel disease (OR = 1.64, 95% CI: 1.30-2.08) and systemic sclerosis (MD of VD: -8.72, 95% CI: -10.11 ~ -7.32) were also found to be statistically significant by MA of observational studies (41, 42).

Cognitive disorders

The association between autism spectrum disorder and vitamin D level was studied in one MA of observational studies. Patients with autism spectrum disorder had a decreased level of 25(OH)D concentration compared to control group (MD: -8.63, 95% CI: -13.17 ~ -4.09, P<0.0001) (45). Cognitive impairment was studied in one MA of observational studies. The association was statistically significant in their summary results (OR = 1.56, 95% CI: 1.05-2.33, *P*=0.028; MD of VD: -6.83, 95% CI: -11.36 ~ -2.30, *P*<0.001) (46).Two MA studies of observational studies summarized the evidence for association between vitamin D level and dementia from previous studies, and both reported statistically significant associations (OR = 1.49, 95% CI: 1.09-1.88; OR = 1.54, 95% CI: 1.19-1.99) (47, 48). Association between vitamin D level and schizophrenia was studied by one MA of observational studies, and the association was statistically significant from their study (MD of VD: -5.91, 95% CI: -10.68 ~ -1.14, cases vs controls; OR = 2.16, 95% CI: 1.32-3.56) (49).

Infectious diseases

Infection of *clostridium difficile* was studied by one MA of observational studies. They found that *clostridium difficile* infection (WMD of VD: -3.54, 95% CI: -6.89 ~ -0.39,

cases vs controls) and *clostridium difficile* infection severity (OR = 1.61, 95% CI: 1.02-2.53) were statistically significantly associated with vitamin D level (50). In addition, mortality among pneumonia patients was also found to be statistically significantly associated with vitamin D level by MA of observational studies (RR = 2.59, 95% CI: 1.32-5.08) (51).

Metabolic disorders

The association between vitamin D concentration and antiphospholipid syndrome was studied in one MA of observational studies, which reported a statistically significant association (SMD of VD: -3.605, 95% CI: -5.449, -1.761, $P<0.001$, cases vs controls; OR of vitamin D deficiency: 3.063, 95% CI: 2.120-4.426, $P<0.001$, cases vs controls) (52).

The associations between vitamin D concentration and diabetic nephropathy was assessed by one study, and it suggested a statistically significant association (OR=1.80, 95% CI: 1.25-2.59, $P=0.002$) (54). Diabetic peripheral neuropathy was assessed as the outcome in two MA studies of observational studies, and both consistently suggested statistically significant associations (OR = 1.22, 95% CI: 1.17-1.27; OR = 2.88, 95% CI: 1.84-4.50) (53, 55). The association between vitamin D and diabetic retinopathy was explored by one MA of observational studies, also suggesting a statistically significant association (OR = 2.03, 95% CI: 1.07-3.86, $P=0.03$) (56).

Three MA of observational studies were on association between obesity or overweight and vitamin D, and all of them consistently reported statistically significant associations (57-59). From MA of RCTs, one study summarizing evidence from studies on obese individuals did not find any statistically significant between vitamin D supplements and BMI or weight (102). Two other MA of RCTs were on fat mass, fat mass percentage and weight in general population, but did not find any statistically significant association (110, 111).

A MA of observational studies was conducted on risk of type 2 diabetes in old adults, and they found a statistically significantly elevated risk of 1.31 (95% CI: 1.11-1.54,

P=0.001) (60). One MA of observational studies was on association between risk of urolithiasis and 1,25(OH)D or 25(OH)D, and 1,25(OH)D level was found to be statistically significantly higher in stone formers compared with controls (MD=10.19, 95% CI: 4.31-16.07, *P*=0.0007) (61).

One MA of RCTs explored the associations between vitamin D supplementations and adiponectin concentration or leptin concentration but did not find any statistically significant change following vitamin D treatment (108). Another MA of RCTs summarized evidence regarding association of vitamin D supplements with fasting plasma glucose, HbA1c and HOMA-IR in T2D patients, all three biomarkers were statistically significantly reduced in intervention group (109). In addition, one MA of RCTs explored the association of vitamin D supplements with HDL-C, LDL-C, HOMA-IR, and triglycerides control in obese individuals, only LDL-C was found to be statistically significantly changed (SMD = 0.338, 95% CI: 0.071-0.605) (102).

Neonatal/infant/child related respiratory outcomes

Two MA of observational studies were on child-related outcomes. One study explored the association between child circulatory vitamin D concentration and lower respiratory tract infection and reported a statistically significant association (OR=3.29, 95% CI: 1.27-8.56, *P*=0.01; MD of VD: -8.75, 95% CI: -15.70 ~ -1.80, *P*=0.01) (63). The other MA was on associations between maternal vitamin D level and respiratory/allergic outcomes in children. They found that maternal vitamin D was statistically significantly associated with risk of respiratory tract infection in children (OR = 0.64, 95% CI: 0.47-0.87, *P*=0.005) (62).

Two MA of RCTs explored the association of vitamin D supplementations with risk of asthma exacerbation in children and reported statistically significant protective effects (RR = 0.41, 95% CI: 0.27-0.63, *P*<0.001; RR = 0.41, 95% CI: 0.27-0.63, *P*<0.001) (112, 114). Another MA of RCTs studied the risk of asthma exacerbation requiring treatment with systemic corticosteroids and found a statistically significant association with vitamin D (RR = 0.74, 95% CI: 0.56-0.97) (113).

Pregnancy related outcomes

Two MA of observational studies were on assisted reproductive outcomes (64, 65). Biochemical pregnancy was studied by one of them and it was statistically significant (OR = 1.34, 95% CI: 1.04-1.73, *P*=0.03). The risk for clinical pregnancy was studied by both, but neither was statistically significant. Live birth was explored by both studies, and both showed a statistically significant protective effect of high vitamin D concentration (OR = 1.33, 95% CI: 1.08-1.65, high VD vs low; RR = 0.75, 95% CI: 0.61-0.93, low vs high). Risk of miscarriage was studied by only one study, but not statistically significant.

The risk of polycystic ovary syndrome was studied by three MA of RCTs, and results from the three study were not consistent. One of them did not find a statistically significant difference in vitamin D concentration between polycystic ovary syndrome patients and controls (SMF of VD: -0.45, 95% CI: -1.68 ~ 0.79) (67). The other two MA of observational studies consistently reported a statistically significant difference (SMD of VD: -0.64, 95% CI: -1.12 ~ -0.15; SMD of VD: -0.86, 95% CI: -1.46 ~ -0.26) (66, 68).

Spontaneous abortion, spontaneous preterm birth and still birth were all assessed for their association with maternal vitamin D status as well, but the study did not find statistically significant association (69).

Skeletal outcomes

Chronic widespread pain was studied in one MA of observational studies, and individuals with chronic widespread pain were at higher risk of vitamin D deficiency compared to controls (OR = 1.63, 95% CI: 1.20-2.23) (70). Similarly, another MA of observational studies explored the association between low back pain and vitamin D concentration, and reported a statistically significant result (OR = 1.60, 95% CI: 1.20-2.12) (72). Another MA of observational studies explored fibromyalgia and found that the 25(OH)D levels of individuals with fibromyalgia were statistically significantly lower than controls (SMD of VD: -0.56, 95% CI: -1.05 ~ -0.08) (71). For MA of RCTs, chronic nonspecific musculoskeletal pain was studied and the effect of vitamin D

supplementation was not statistically significant (SMD: 0.05, 95% CI: -0.37~0.46) (92). In another MA, vitamin D supplementation was found to statistically significantly reduce pain score (MD: -0.57, 95% CI: -1.00 ~ -0.15) (96). In addition, vitamin D supplementation was also found to have statistically significant effect on visual analog scale of pain intensity among chronic widespread pain patients (MD=0.46, 95 CI: 0.09-0.89, *P*=0.02) (97).

The association between vitamin D level and walking speed in elderly was studied in a MA of observational studies and they found a statistically significant association (MD = -0.08, 95 CI%: -0.09 ~ -0.07, P<0.001, vitamin D deficiency vs normal vitamin D profile) (74). There was a MA of RCTs summarizing the association of vitamin D supplementation with physical performance outcomes, including hand grip strength, leg strength, physical performance, time up and go (a measurement for physical performance, which is the time that a person takes to rise from a chair, walk three meters, turn around, walk back to the chair) and walking capacity in old adults. Vitamin D supplement was found to nominally shorten time up and go (MD = -0.75, 95% CI: -1.44 ~ -0.07, *P*=0.03), and none of the other outcomes studied were statistically significant.

There was a MA of observational studies on association between vitamin D level and stress fracture in a military population which found that compared with controls, cases had a statistically significantly lower 25(OH)D level (MD: -2.44, 95% CI: -4.05 ~ -0.84, *P*=0.003) (73).

Two MA of RCTs focused on the benefits of vitamin D supplementation on knee osteoarthritis. Western Ontario and McMaster Universities Arthritis Index (WMAC) function and WOMAC pain was found to be improved by vitamin D supplements. However, these two study did not support any effect of vitamin D supplements on joint space width and WOMAC stiffness (94, 95).

Other outcomes

The association between vitamin D level and non-cardiovascular, non-cancer death (death due to any reasons other than cardiovascular events or cancer) was studies by a MA of observational studies. Compared to the top third of the vitamin D level distribution, individuals in the bottom third were 34% increased risk of non-cardiovascular, non-cancer death (RR = 1.34, 95% CI: 1.13-1.60) (86). A MA study of RCTs explored the effect of vitamin D supplementation on all-cause mortality in CKD patients, and did not find any statistically significant effect (RR = 0.84, 95% CI: 0.46-1.52) (101).

Two MA studies of observational studies were on association between vitamin D and outcomes in critically ill patients (patients in intensive care unit). One study explored 30-day mortality of patients in ICU, and found a nominal statistically significant risk for patients with low vitamin D level (RR = 1.42, 95% CI: 1.00-2.02, $P = 0.05$) (75). Another study explored mortality and suggested a statistically significant elevated risk (OR = 1.76, 95% CI: 1.38-2.23) (84). Infection, in-hospital mortality and sepsis among patients in ICU were also included as outcomes in the first study, and the results were statistically significant for all of the three outcomes (75). Finally, another MA of observational studies explored the association between vitamin D level and sepsis in hospitalized patients, and found a statistically significant elevated risk for patients with low vitamin D level as well (OR = 1.78, 95% CI: 1.55-2.03, $P<0.001$) (91). Two MA of RCTs involved length of hospital stay, length of ICU stay, length of mechanical ventilation and mortality among ICU patients as their outcomes. Results from the two studies were consistent, and only mortality risk was found to be statistically significantly reduced by vitamin D supplements (RR = 0.84, 95% CI: 0.66-1.06; OR = 0.70, 95% CI: 0.50-0.98) (118, 119).

Two MA of observational studies were on non-alcoholic fatty liver disease (NAFLD). The NAFLD fibrosis score and activity score were not statistically significantly associated with vitamin D level (MD: 0.88, 95% CI: -2.65 ~ 4.42 for fibrosis score; MD: -0.93, 95% CI: -2.45 ~ 0.58 for activity score) (82). However, the risk of NAFLD was found to be statistically significantly associated with vitamin D level in another

MA (OR=1.26, 95% CI: 1.15-1.38), and the difference of vitamin D level between non-alcoholic steatohepatitis patients vs controls was also explored and reported to be statistically significant by this study (SMD of VD: -1.30, 95% CI: -2.37 ~ -0.23) (85). Two MA of observational studies were on age-related macular degeneration. The results from the two studies were inconsistent. One of them reported a nominally statistically significant association when the highest quintile of vitamin D level was compared with the lowest quintile (SMD of VD: -0.15, 95% CI: -0.41 ~0.11, *P*=0.272, cases vs controls; OR = 0.83, 95% CI: 0.71-0.97, *P*=0.019, highest quintile vs lowest quintile of vitamin level) (78). However, the other MA did not find any statistically significant association (OR = 0.91, 95% CI: 0.69-1.22, *P*=0.12, for an increase of 25(OH)D concentration by 10 ng/mL), partly due to the different comparison used from the previous study (121).

Two MA of observational studies were on chronic obstructive pulmonary disease (COPD) and both found a statistically significantly higher vitamin D concentration in controls compared to cases with COPD (SMD of VD: 0.60, 95% CI: 0.31 ~ 0.89; SMD of VD: 0.69, 95% CI: 0.38 ~ 1.00) (80, 81).

Three MA of observational studies were on risk of Parkinson's disease, and all of them suggested a statistically significant association between vitamin D and risk of Parkinson's disease. Compared with controls, cases with Parkinson's disease had a lower 25(OH)D concentration (SMD of VD: -16.88, 95% CI: -33.54 ~ -0.23; MD of VD: -11.55, 95% CI: -12.23 ~ -10.86) (88, 89). In individuals with low vitamin D level, the risk of Parkinson's disease was statistically significantly elevated compared to individuals with high vitamin D level (OR = 1.50, 95% CI: 1.14-1.97; OR = 1.50, 95% CI: 1.31-1.71) (88, 90).

1.2.3 Association between vitamin D and health outcomes – published review

From the published umbrella review in 2014, a total of 137 outcomes were explored in their relationship with vitamin D. These included skeletal outcomes, autoimmune diseases, cancer outcomes, cardiovascular outcomes, cognitive disorders, infectious diseases, metabolic disorders, neonatal/infant/child related outcomes, pregnancy related outcomes, and some other outcomes. In this section, I will give more details for outcomes of high prevalence/high burden based on the articles which were included in the published umbrella review of vitamin D. For more information, please refer to the published version of the vitamin D umbrella review (25) .

1.2.3.1 Skeletal outcomes

As has already been described above, vitamin D regulates calcium and phosphorus homeostasis. In a systematic review including 167 studies some evidence of an association between low 25(OH)D and established rickets was found (122). In addition, while evidence on impact of 25(OH)D on bone mineral contents in infants was found to be inconsistent, associations with bone mineral density in adolescents and the elderly were identified (122). However, in contrast to these findings, a systematic review and meta-analysis by Winzenberg and colleagues, including 6 studies in children and adolescents (343 participants receiving placebo vs 541 receiving vitamin D), reported a weak trend showing a small increase in lumbar spine bone mineral density with increased vitamin D concentration (SMD: 0.15, 95% CI: -0.01 ~ 0.31, P = 0.07) (123). Another systematic review and meta-analyses of RCTs published in 2014 included 23 studies on bone mineral density, and only a small benefit at the femoral neck was found, with no effects on any other sites (124).

A systematic review and meta-analysis of 17 case-control studies found 33% lower serum 25(OH)D levels in cases of hip fracture compared to controls (1903 fracture cases vs 1953 controls), however, the pooled result of RCTs did not support any effect (8 RCTs; pooled RR = 1.13, 95% CI: 0.98-1.29) (125). Furthermore, a statistically significant protective effect of vitamin D supplementation against non-vertebral was identified in meta-analysis of 12 RCT studies (for non-vertebral fracture: 12 RCTs,

pooled RR = 0.86, 95% CI: 0.77-0.96; for hip fracture: 8 RCTs, pooled RR = 0.91, 95% CI: 0.78-1.05), which did not suggest any association between vitamin D supplements and risk of hip fracture (126). However, robust evidence was still lacking regarding whether vitamin D supplementation alone is effective in osteoporosis and fracture prevention. In 2010, a meta-analysis of 7 major RCTs including 68517 participants, DIPART (Vitamin D Individual Patient Analysis of Randomized Trials) Group, concluded that vitamin D alone was not effective, while given together with calcium, it could reduce risk of hip fractures and total fractures and probably vertebral fractures (127). In contrast, a more recent meta-analysis including 31022 participants found that vitamin D supplementation of more than 800 IU per day reduced the risk of hip fracture and any non-vertebral fracture independent of additional calcium intake (For hip fracture: HR = 0.70, 95% CI: 0.58-0.86; For non-vertebral fracture: HR = 0.86, 95% CI: 0.76-0.96) (128). Notably, agreement that combining vitamin D and calcium supplementation together could reduce fracture risk was reached by many reviews (129-133).

For risk of falls, a meta-analysis of RCTs found that fall risk was statistically significantly reduced by high dose supplemental vitamin D (700-1000 IU/day) (RR = 0.81, 95% CI: 0.71-0.92 from 7 RCTs involving 1921 participants), and meanwhile achieving serum 25(OH)D concentrations of 60 nmol/L or more reduced fall risk by 23% (RR = 0.77, 95% CI: 0.65-0.90) (134). Another meta-analysis including 5 RCTs found that the OR of falling was reduced by 22% by vitamin D (OR = 0.78, 95% CI: 0.64-0.92, 5 RCTs involving 1237 participants) (135).

1.2.3.2 Cancer

In a meta-analysis including 8 prospective studies, the top quantile of 25(OH)D had a OR of 0.66 (95% CI: 0.54-0.81) compared with bottom quantile on colorectal cancer risk, and for rectal cancer (OR=0.50, 95% CI: 0.28-0.88) the protective effect was stronger than that of colon cancer (OR=0.77, 95% CI: 0.56-1.07) (136). In addition, the association between 25(OH)D and colorectal adenoma incidence was also identified (OR=0.82, 95% CI: 0.69-0.97) by a meta-analysis of 10 observational studies (137).

Evidence on associations between vitamin D levels and prostate cancer was inconclusive from meta-analyses studies of observational studies (138-140). In particular, in a systematic review and meta-analyses study including 25 papers published in 2011, the OR (95% CI) for per 1000 IU increase in dietary intake, per 10 ng/mL increase in serum 25(OH)D concentration, per 10 pg/mL increase in 1,25(OH)D concentration on total prostate cancer was 1.14 (0.99-1.31), 1.04 (0.99-1.10) and 1.00 (0.87-1.14), while on aggressive prostate cancer it was 0.93 (0.63-1.39), 0.98 (0.84-1.15) and 0.86 (0.72-1.02) respectively (139).

In respect of breast cancer, a meta-analysis of 21 observational studies reported that women in the highest quantile of 25(OH)D were found to have a reduced risk of breast cancer risk of 0.86 (95% CI: 0.75-1.00) compared with those in the lowest quantile in nested case-control and retrospective studies, and the risk was 0.35 (95% CI: 0.24-0.52) from meta-analysis of case-control studies (141). In a meta-analysis of 9 prospective studies, an inverse association was found between serum 25(OH)D (27-35 ng/mL) and breast cancer risk in postmenopausal women, with every 5 ng/mL increase associated with a 12% reduced risk (RR=0.88 per 5ng/mL, 95% CI: 0.79-0.97) (142). However, other meta-analyses failed to replicate these findings (132, 138). In a meta-analysis including 4 nested case-control studies (2363 cases vs 2363 controls), it was shown that each 10 nmol/L increase in 25(OH)D level was not associated with risk for breast cancer (OR = 0.99, 95% CI: 0.97-1.01, *P*=0.42) (132). In another meta-analysis of 10 observational studies, although an overall association between vitamin D and breast cancer was shown (RR = 0.89, 95% CI: 0.81-0.98), it only existed in case-controls studies (5 studies of 3030 cases, pooled RR = 0.83, 95% CI: 0.79-0.87). In 5 prospective studies with 3145 cases, the pooled RR was 0.97 (95% CI: 0.92-1.03) (138).

1.2.3.3 Cardiovascular diseases

In a meta-analysis including 18 observational studies, the pooled OR of hypertension was 0.73 (95% CI: 0.63-0.84) for the highest category of serum 25(OH)D level compared to the lowest (143). In meta-analyses of RCTs, small effects of vitamin D

supplementation on both systolic (-2.44 mm Hg, 95% CI: -4.86 to -0.02) and diastolic (-3.1 mm Hg, 95% CI: -5.5 to -0.6) blood pressure were found (143-145).

Various cardiovascular outcomes have been found to be associated with vitamin D status. In a meta-analysis of ten observational studies (58384 participants, and 2644 cases), the highest quartile of 25(OH)D level had a statistically significant elevated risk for ischemic stroke compared to individuals in the lowest quartile (OR = 1.54, 95% CI: 1.43-1.65) (146). Similarly, in another meta-analyses including 7 prospective studies (1214 cases with stroke), the pooled risk of incident stroke was reported to be 1.52 (95% CI: 1.20-1.85) (147). Another meta-analysis of prospective studies (24 studies included, involving 6123 CVD cases and 65994 total participants) reported a pooled RR of 1.64 (95% CI: 1.27-2.10) for stroke, and they also reported statistically significantly elevated risk for total CVD, CVD mortality and coronary heart disease (total CVD: RR = 1.52, 95% CI: 1.30-1.77; CVD mortality: RR = 1.42, 95% CI: 1.19-1.71; coronary heart disease: RR = 1.38, 95% CI: 1.21-1.57) (148). A meta-analysis of observational studies reported that risk for ischemic heart disease was increased by 39% (HR = 1.39, 95% CI: 1.25-1.54, 18 studies included), and risk for early death was increased by 46% (HR = 1.46, 95% CI: 1.31-1.64, 17 studies included) (149).However other meta-analyses failed to replicate these findings. In a meta-analysis including 50 RCTs with 294,478 participants, no effects of supplementation with vitamins and antioxidants was found for any major cardiovascular events. Specifically, for vitamin D, including 7 studies, the summary RR was 1.02 (95% CI: 0.98-1.07) for prevention of major cardiovascular events (150).

1.2.3.4 Metabolic disorders and diabetes

Forouhi and colleagues incorporated their own data in a meta-analysis of prospective studies (11 studies included, 3612 cases and 55713 controls), and found that the combined RR of type 2 diabetes risk comparing the highest with lowest quartile of 25(OH)D was 0.59 (95% CI: 0.52-0.67) (151). Similarly, in a study of 9841 participants and a subsequent meta-analysis (14 studies included, 4877 cases and 72204 total participants), Afzal and colleagues reported an association between low serum 25(OH)D with increased risk of type 2 diabetes (OR=1.50, 95% CI: 1.33-1.70)

(152). However other meta-analyses and/ or systematic reviews failed to replicate these findings. A meta-analysis of RCTs showed a small effect on fasting glucose and small improvement in insulin resistance among patients with diabetes or impaired glucose tolerance, and the authors concluded that evidence was insufficient regarding the beneficial effect of vitamin D supplementation (153).

1.2.3.5 Cognitive Disorders

Relationships between vitamin D and cognitive performance and disorders have also been recently examined. In a systematic review and meta-analysis, Balion and colleagues showed that the group with higher vitamin D levels (>= 50 nmol/L) had a higher Mini-Mental State Examination score (WND = 1.2, 95% CI: 0.5-1.9, 8 studies included), and in addition, Alzheimer disease cases had a lower vitamin D concentration compared with controls (WMD = -6.2 nmol/L, 95% CI: -10.6 ~ -1.8, 4 studies included) (154). Another meta-analysis in the same year reported similar results, with individuals of low vitamin D levels having an increased risk of cognitive impairment compared with individuals of normal vitamin D level (OR = 2.39, 95% CI: 1.91-3.00, 7 studies included compromising 7688 participants) (155). In addition, vitamin D has also been found to be association with depression. In a meta-analysis of 14 observational studies compromising 31424 participants, individuals with depression were found to have statistically significantly lower vitamin D concentration compared with controls (SMD = -0.60, 95% CI: -0.97 ~ -0.23) (156).

1.3 Vitamin D GWAS

By searching the GWAS Catalogue (https://www.ebi.ac.uk/gwas/), there were three previous GWAS studies on circulating 25-hydroxyvitamin D levels (157-159).

In a European population of 4501 persons, Ahn J and colleagues conducted a GWAS for 25(OH)D concentration and found three loci in statistically significant association with vitamin D (*GC*, *NADSYN1/DHCR7* and *CYP2R1*) (157). The Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits (SUNLIGHT) consortium conducted a GWAS in 33,996 white individuals in year 2010 and found four loci in statistically significant association with vitamin D levels (*GC*, *DHCR7*, *CYP2R1* and *CYP24A1*) (159). In year 2018, as an update of their old GWAS in 2010, SUNLIGHT published a new GWAS in 79,366 individuals, yielding two new loci in association with vitamin D (*AMDHD1* and *SEC23A*) (158). For the details of genes and top variants identified by previous GWAS, please see **Table 5**.

Table 5. Genes associated with 25-hydroxyvitamin D level by previous GWAS.

Study	Variant	Gene	Chromosome: Position	*P* value
Ahn J.	rs2282679	*GC*	4: 71742666	$1.80*10^{-49}$
Wang T.J.	rs2282679	*GC*	4: 71742666	$1.90*10^{-109}$
Jiang X.	rs3755967	*GC*	4: 71743681	$4.74*10^{-343}$
Jiang X.	rs10741657	*CYP2R1*	11: 14893332	$2.05*10^{-46}$
Wang T.J.	rs10741657	*CYP2R1*	11: 14893332	$3.30*10^{-20}$
Ahn J.	rs2060793	*CYP2R1*	11: 14893764	$2.90*10^{-17}$
Jiang X.	rs12785878	*NADSYN1/DHCR7*	11: 71456403	$3.80*10^{-62}$
Wang T.J.	rs12785878	*NADSYN1/DHCR7*	11: 71456403	$2.10*10^{-27}$
Ahn J.	rs3829251	*NADSYN1/DHCR7*	11: 71483513	$3.40*10^{-9}$
Jiang X.	rs10745742	*AMDHD1*	12:95964751	$2.10*10^{-20}$
Jiang X.	rs8018720	*SEC23A*	14:39086981	$1.11*10^{-11}$
Wang T.J.	rs6013897	*CYP24A1*	20: 54125940	$6.00*10^{-10}$
Jiang X.	rs17216707	*CYP24A1*	20:54115823	$8.14*10^{-23}$

Note: "chromosome: position" used version GRCh38.p7.

The gene *GC* encodes vitamin D binding protein (DBP), which is a protein compromising 474 amino acids (160). DBP are found in multiple sites, including plasma, ascetic fluid, cerebrospinal fluid and on the surface of multiple cell types. In relation to vitamin D, it binds with vitamin D in skin and transports vitamin D to liver for 25 hydroxylation (161).

The gene *CYP2R1* (cytochrome P450 family 2, Subfamily R, Member 1) encodes an enzyme belonging to the cytochrome P450 superfamily. This enzyme catalyses many reactions involved in the synthesis of cholesterol, steroids and other lipids (162) and it could be the enzyme underlying 25-hydroxylation of vitamin D in the liver (162).

The gene *DHCR7* (7-Dehydrocholesterol Reductase) encodes an enzyme that removes the double bond in the B ring of sterols and catalyses the conversion of 7-DHC to cholesterol which is a vital step in the synthesis of vitamin D (163).

The gene *AMDHD1* (Amidohydrolase Domain Containing 1) is a protein coding gene, which encodes a protein related to histidine degradation and viral mRNA translation (164).

The protein encoded by gene *SEC23A* (Sec 23 Homolog A, Coat Complex II (COPII) Component) is a member of the SEC23 subfamily. In eukaryotic cells, secreted proteins are synthesized in the endoplasmic reticulum, wrapped by COPII-coated vesicles, and subsequently transported to the Golgi apparatus. SEC23 protein plays a role in promoting endoplasmic reticulum – Golgi protein transportation as part of COPII complex (158).

The gene *CYP24A1* (Cytochrome P450 Family 24 Subfamily A Member 1) encodes a protein belonging to the cytochrome P450 superfamily. It catalyses the inactivation of $1,25(OH)_2D$ by hydroxylation of the side chain on C24, which converts $1,25(OH)_2D$ to calcitriol acid (165).

1.4 Systematic literature review of vitamin D Mendelian Randomization Studies

Mendelian Randomization (MR) analysis employs exposure (serum 25(OH)D level in my project) related SNPs as instruments and explores the causal relationship between exposure and outcomes by analysing the association between instruments and outcomes. In a MR study, G denotes the instrumental variable (e.g. genetic variant(s)), E denotes the exposure of interest, and O denotes the outcomes/phenotypes (e.g. disease). The direct association between E and O may be biased by measured/unmeasured confounders (denotes by U). If G is associated with E and free from U, the causal effect of E on O can be estimated by the relationship between G and O (**Figure 6**).

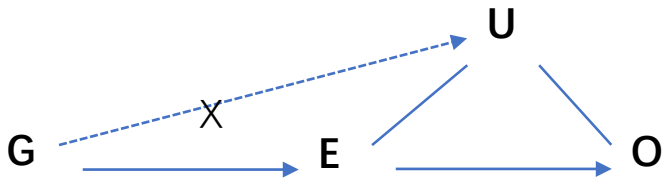


Figure 6. Rationale of a Mendelian Randomisation Study.

MR can, therefore, overcome the limitation of confounding factors and reverse causality. The idea of IV was first introduced in the econometrics literature by Wright and later adopted by the statistical measurement error and causal inference literature (166, 167). There are three assumptions underlying MR: 1) IV is associated with exposure (e.g., serum 25(OHD)); 2) IV is independent of outcome conditional on exposure and confounders (no pleiotropy); 3) IV is not associated with confounders (168). Genetic markers are therefore excellent IVs since they are randomly allocated at inception and thus randomly distributed (Mendel's second law). This means that, even at population level, when relating genetic variants to diseases, alleles are generally unrelated to confounders and the association between genotypes and diseases are protected from reverse causality (169).

In collaboration with Yazhou He, another PhD student in Usher Institute, University

of Edinburgh, I conducted a systematic review for all published Mendelian Randomisation studies on vitamin D. Yazhou He ran the literature search. We both did the screening of the studies and I did the data extraction.

1.4.1 Search strategy and review process

For the systematic literature search, Medline and Embase were accessed in February 2017. The search strategy is presented in **Table 6**. I included articles which were MR studies on association between vitamin D and health outcomes. Any review article, non-English article, letter or conference abstract was excluded. Studies on the impact of other markers/exposures on vitamin D levels were also excluded. After deletion of duplications, 95 references went into subsequent screening. Since the number of references is relative small, I downloaded full texts for all of them and implemented full text screening. After full text review, 27 studies were eligible for inclusion. At last, 2 more eligible studies were identified by tracking the reference lists of the 27 included studies. As a result, a total of 29 studies met our inclusion criteria (**Figure 7**).

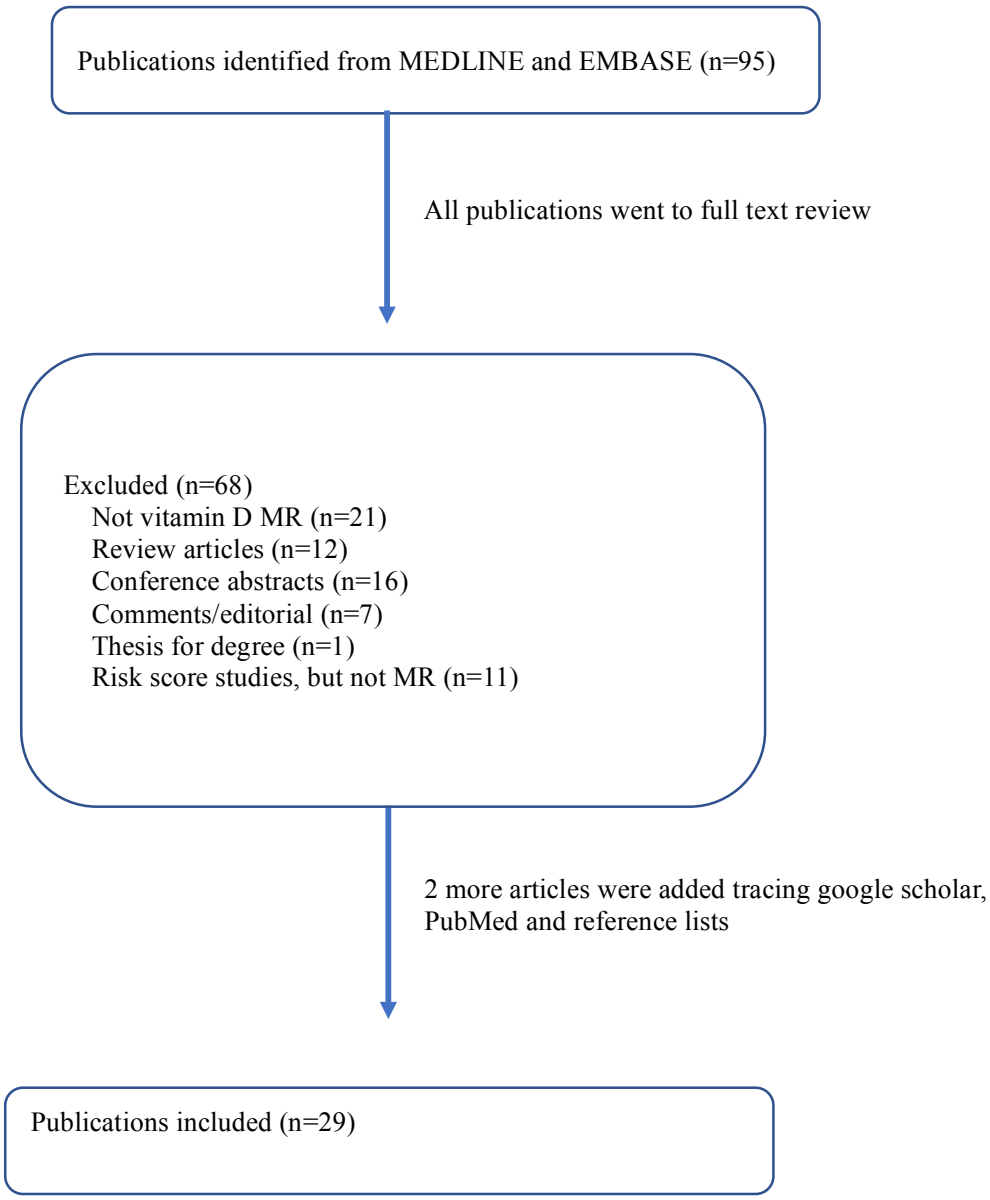
Table 6. Search strategy and algorithm used in the systematic review for vitamin D Mendelian Randomization studies

MEDLINE (OvidSP)
1. Vitamin d/ or 25-OHD.mp. or 25 hydroxyvitamin D.mp. or cholecalciferol/
2. colecalciferol.mp. or hydroxycholecalciferols/ or hydroxycolecalciferols.mp.
3. calcifediol/ or dihydroxycholecalciferols/ or dihydroxycolecalciferols.mp.
4. calcitriol/ or 24,25-dihydroxyvitamin d 3/ or 24,25-OH2 D3.mp.
5. ergocalciferols/ or dihydrotachysterol/ or 25-hydroxyvitamin d 2/ or 25-OHD2.mp.
6. 1,25-dihydroxyvitamin d.mp. or 1,25-OH2 D.mp. or 1,25-dihydroxyvitamin d2.mp.
7. 1,25-dihydroxyergocalciferol.mp. or 1,25-OH2D2.mp. or 1,25-dihydroxyvitamin d3.mp. or 1,25-OH2 D3.mp. or ergocalciferols/
8. vitamin D2.mp. or vitamin D 2.mp. or vitamin D3.mp. or vitamin D 3.mp
9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10. Mendelian Randomization Analysis/ or Mendelian randomisation.mp or Mendelian randomization.mp
11. instrumental variable.mp or instrumental variables.mp or genetic instrument.mp or genetic instruments.mp
12. random Mendelian.mp or genetic risk score.mp or genetic risk scores.mp or genetic score.mp or genetic scores.mp
13. 10 or 11 or 12
14. 9 and 13
EMBASE (OvidSP)
1. Vitamin d/ or 25-OHD.mp. or 25 hydroxyvitamin D/ or colecalciferol/ or cholecalciferol.mp.
2. hydroxycolecalciferols/ or hydroxycholecalciferols.mp.
3. calcifediol/ or dihydroxycolecalciferols/ or dihydroxycholecalciferols.mp.
4. calcitriol/ or secalciferol/ or 24,25-OH2 D3.mp. or ergocalciferol/ or dihydrotachysterol/
5. 25-hydroxyvitamin d 2.mp. or 25-OHD2.mp. or 1,25-dihydroxyvitamin d.mp. or 1,25-OH2 D.mp.
6. 1,25dihydroxyergocalciferol/ or 1,25-dihydroxyvitamin d2.mp. or 1,25-OH2 D2.mp.
7. 1,25-dihydroxyvitamin d3.mp. or 1,25-OH2 D3.mp. or ergocalciferol derivative/
8. vitamin D2.mp. or vitamin D 2.mp. or vitamin D3.mp. or vitamin D 3.mp
9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10. Mendelian Randomization Analysis/ or Mendelian randomisation.mp or Mendelian randomization.mp

11. instrumental variable.mp or instrumental variables.mp or genetic instrument.mp or genetic instruments.mp
12. random Mendelian.mp or genetic risk score.mp or genetic risk scores.mp or genetic score.mp or genetic scores.mp
13. 10 or 11 or 12
14. 9 and 13

mp: any match for title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword.

Figure 7. Flowchart for vitamin D MR systematic review process



1.4.2 Result summary

For a summary of the most highly powered or most recent study for each outcome among the 29 eligible studies, please see **Table 7**.

Among the 29 included studies, 5 were on glycaemic traits and diabetes, two of which found suggestive causal associations (170-174). Afzal et al. conducted MR for the effect of 25OHD on T2D in 96,423 white Danish individuals (5037 T2D cases). The genes *DHCR7* and *CYP2R1* were employed as IVs separately. For a sum score of 2 SNPs (rs11234027, rs7944926) in *DHCR7*, the OR was estimated to be 1.51 (95% CI: 0.98-2.33, *P*=0.04) per 20 nmol/L lower 25OHD as determined by genetic factors, which provided nominal evidence suggesting a causal relationship between vitamin D and T2D. However, in their own study, the OR for the *CYP2R1* score (sum score of variants rs10741657 and rs12794714) was 1.02 (95%CI: 0.75-1.37, *P*=0.84). The inconsistent results might be due to *DHCR7* and *CYP2R1* being involved in distinctive steps of vitamin D synthesis, and only the step *DHCR7* involved in (involved in skin synthesis of vitamin D_3 from sun exposure) being associated with increased risk of T2D (171). In addition, Husemoen et al. studied the association between vitamin D and adiponectin, which is a protein involved in several metabolic processes, including glucose regulation and had been reported to be statistically associated with T2D (172). In a population of 6405 Danish individuals, they found that a doubling of 25(OH)D level was causally associated with a 61.46% (95% CI: 17.51-120.28, *P*=0.003) higher adiponectin. None of the other 3 studies found any evidence supporting a causal role of vitamin D in the glycaemic pathway or on risk of diabetes. Notably, Ye et al. conducted a summary statistics MR in populations of European descent with data from several consortiums (EPIC-InterAct, DIAGRAM, ADDITION-Ely, Norfolk Diabetes and Cambridgeshire) with SNPs in four genes reported to be associated with serum 25OHD level by SUNLIGHT Consortium (159) as IVs, and for T2D they reached 28,144 cases and 76,344 controls. Their estimated OR for risk of T2D was 1.01 (95% CI: 0.75-1.36, *P*=0.94) per SD decrease in 25OHD level as determined by genetic factors. In addition, they also analysed glycaemic traits including fasting glucose, 2-hour glucose, fasting insulin, HbA1c, but none of them were statistically significantly

associated with vitamin D (174).

Concerned by its impact on circulatory outcomes, a total of 5 previous studies were on blood pressure and circulatory diseases (175-179). Skaaby and colleagues used three common filaggrin gene mutations of *R501X*, *2282del4* and *R2447X* as IVs and studied their associations with blood pressure, cholesterol, BMI, waist circumference and metabolic syndrome. Their analysis suggested a 23.8% (95% CI: 3.0-48.6, *P*=0.02) increase in HDL-cholesterol level and 30.5% (95% CI: 0.8-51.3, *P*=0.04) lower triglycerides level per doubling of vitamin D level (178). Another study employed a synthesis score of genes *DHCR7* and *CYP2R1*, and with multiple consortia data of up to 140k individuals, each 10% increase in vitamin D levels was found to be associated with 0.37 mmHg (95% CI: -0.003-0.73, *P*=0.052) lower systolic blood pressure, 0.29 mmHg (95% CI: 0.07-0.52, *P*=0.01) lower diastolic blood pressure and 8.1% decreased risk of hypertension (OR=0.92, 95% CI: 0.87-0.97, *P*=0.002) (179). The other 3 studies did not find any statistically significant associations.

There were 4 studies that focused on autoimmune diseases (180-183). In a consortium of 14,498 multiple sclerosis (MS) cases and 24,091 controls, Morkry et al. found that the risk for MS was elevated by 1.02 (OR=2.02, 95% CI: 1.65-2.46, $P=7.72*10^{-12}$) per standard deviation decrease in log-transformed 25OHD level as determined by IVs (4 vitamin D related loci as previously identified by SUNLIGHT Consortium in 2010 (159)) (180). Similarly, using 3 SNPs as IVs (rs2282679 in *GC*, rs2060793 in *CYP2R1* and rs3829251 in *DHCR7*), Rhead et al. observed an OR of 0.85 (95% CI: 0.76-0.94, *P*=0.003) in a white population of 7,391 MS cases and 14,777 controls (181). The other 2 studies were on rheumatoid arthritis severity and response to anti-tumour necrosis therapy for RA patients, but the outcomes they tested were not statistically significantly associated with the vitamin D IVs.

Three studies explored cognitive functions and related outcomes (184-186). Jorde et al. tested the association between SNPs related to 25(OH)D level and cognitive functions, including word call test, digit-symbol coding test and finger tapping test. Only rs7975232 in gene *VDR* showed an association with word call test and digit-

symbol coding test (homozygote of minor allele had significantly higher average mark than major allele homozygote). They tested 5 SNPs in *VDR*, one SNP in each of *NADSYN1*, *CYP2R1*, *CYP24A1*, *DBP* and a risk score, however, none of the others showed any statistically significant association (184). In addition, in a prospective cohort of 1207 white individuals, Kueider et al. investigated the causal effect of vitamin D on cognitive functions during aging with two SNPs (rs2282679 and rs7041) in *GC* gene as IVs. The score of the two SNPs was found to be causally associated with clock drawing task and psychomotor speed (185). Finally, Morkry studied the effect of vitamin D on risk of Alzheimer disease (AD) within the International Genomics of Alzheimer's Project (17,008 cases and 37,154 controls). Meta-analysis of the associations between 4 SNPs (reported by SUNLIGHT Consortium in 2010 (159)) and AD showed a 25% elevated risk (OR=1.25, 95% CI: 1.03-1.51, *P*=0.021) per standard deviation decrease of log-transformed 25OHD level (186).

There were 3 studies on mortality or survival outcomes (187-189). Afzal et al. used four SNPs (rs11234027 and rs79449256 in *DHCR7*; rs10741657 and rs12794714 in *CYP2R1*) as IVs and studied their association with all-cause mortality, cardiovascular mortality, cancer morality, and mortality from other causes in three Danish cohorts of 95,766 individuals and 10,349 deaths. From their IV analysis, every 20nmol/L decreased 25(OH)D level determined by genetic variants conferred a 30% increased risk to all-cause mortality (10,349 deaths; OR=1.3, 95% CI: 1.05-1.61), a 43% increased risk on cancer mortality (2839 deaths; OR=1.43, 95% CI: 1.02-1.99) and a 44% increased risk on other mortality (2585 deaths; OR=1.44, 95% CI: 1.01-2.04). The association between IVs and cardiovascular mortality was not statistically significant (3231 deaths; OR=0.77, 95% CI: 0.55-1.08) (187). Trummer et al. studied the association between 3 25(OH)D related SNPs (rs2282679 in *GC*, rs10741657 in *CYP2R1* and rs12785878 in *DHCR7*) and mortality (including all-cause mortality of 995 deaths, cardiovascular mortality of 619 deaths and non-cardiovascular mortality of 355 deaths) in a prospective study of 3,316 participants. None of the genotypes was found to be associated with mortality outcomes (189). In addition, there was a study on melanoma survival by Davies et al. They studied the association between rs2282679 in *GC*, which has been reported to be associated with vitamin D level, and

overall survival and melanoma-specific survival in a meta-analysis of 3,137 melanoma patients. The SNP was not associated with overall survival but was associated with melanoma specific survival in cohorts where data were available (HR=1.22, 95% CI: 1.04-1.43, *P*=0.01) (188).

There were two studies on cancer (190). Ong et al. explored the effect of vitamin D on ovarian cancer in a sample of 10,065 cases and 21654 controls exploiting 3 SNPs as IVs (rs7944926 in *DHCR7*, rs12794714 in *CYP2R1* and rs2282679 in *GC*). The odds were 1.27 (95% CI: 1.06-1.51) for all ovarian cancer (10,065 cases) and 1.54 (95% CI: 1.19-2.01) for high grade serous type (4121 cases) (191). In addition, Theodoratou et al. studied the impact of vitamin D on colorectal cancer with 4 SNPs (reported by SUNLIGHT Consortium in 2010 (159)) as IVs, but did not find any statistically significant association (190).

One study was on childhood caries. In this study, three SNPs (rs10741657 in *CYP2R1*, rs7944926 in *DHCR7* and rs2282679 in *GC*) were employed as IVs. In a population of 5,545 children of European descent, they did not observe any statistically significant association with caries experience, dental general anesthetic or early caries onset (192). Another study was on bone mineral density and bone metabolism biomarkers (parathyroid hormone and procollagen type 1 N-terminal) in 1824 Chinese postmenopausal women. And they did not find any statistically significant association (193).

There were two studies on body size related outcome. One of them was on birth weight (194) and the other one was on BMI (195). None of them found any significant causal association. Finally, there was one study on paediatric asthma (196), one study on C-reactive protein (197), and one study on atherogenic lipoproteins (198), but none of them found any statistically significant association in their MR analyses.

Table 7. Findings from most highly powered or recent Mendelian Studies on each type of outcome.

Study	Outcomes	Population	No/No of Events	Estimate of effect (95% CI)	*P* value	Metric
Ye Z. 2015 (174)	T2D	white	104488/28144	1.01 (0.75, 1.36)	0.94	1 SD decrease in 25OHD level
Ye Z. 2015 (174)	Fasting glucoes	white	46368	-0.02 (-0.04, 0.01)	0.28	mmol/L per SD decrease in 25OHD level
Ye Z. 2015 (174)	2-h glucose	white	46368	0.08 (-0.06, 0.22)	0.25	mmol/L per SD decrease in 25OHD level
Ye Z. 2015 (174)	Fasting insulin	white	46368	-1.04 (-3.91, 1.83)	0.48	% difference per SD decrease in 25OHD level
Ye Z. 2015 (174)	HbA1c	white	46368	0.01 (-0.04, 0.05)	0.8	% difference per SD decrease in 25OHD level
Vim KS. 2014 (179)	SBP	white	146581	-0.37 (-0.73, 0.003)	0.052	mm Hg per 10% increase in 25OHD level
Vim KS. 2014 (179)	DBP	white	142255	-0.29 (-0.52, -0.07)	0.01	mm Hg per 10% increase in 25OHD level
Vim KS. 2014 (179)	Risk of hypertension	white	142255	0.92 (0.87, 0.97)	0.002	per 10% increase in 25OHD level
Manousaki D. 2016 (177)	coronary artery disease	white	86995/22233	0.99 (0.84, 1.17)	0.93	1 SD decrease in log-transfromed 25OHD level
Morkry LE. 2015 (180)	multiple sclerosis	white	38589/14498	2.02 (1.65, 2.46)	7.72E-12	1 SD decrease in log-transfromed 25OHD level
Morkry LE. 2016 (186)	Alzheimer disease	white	54162/17008	1.25 (1.03, 1.51)	0.021	1 SD decrease in log-transfromed 25OHD level
Afzal S. 2014 (187)	All-cause mortality	white Danish	95766/10349	1.3 (1.05, 1.61)	NA	20 nmol/L lower 25OHD
Afzal S. 2014 (187)	cardiovascular mortality	white Danish	95766/3231	0.77 (0.55, 1.08)	NA	20 nmol/L lower 25OHD
Afzal S. 2014 (187)	cancer mortality	white Danish	95766/2839	1.43 (1.02, 1.99)	NA	20 nmol/L lower 25OHD

Afzal S. 2014 (187)	other mortality	white Danish	95766/2585	1.44 (1.01, 2.04)	NA	20 nmol/L lower 25OHD
Ong JS. 2016 (191)	all ovarian cancer	white	31719/10065	1.27 (1.06, 1.51)	NA	20 nmol/L lower 25OHD
Ong JS. 2016 (191)	high grade serous subtype	white	31719/4121	1.54 (1.19, 2.01)	NA	20 nmol/L lower 25OHD
Theodoratou E. 2012 (190)	colotectal cancer	white	4238/2001	1.16 (0.60, 2.23)	NA	per unit increase in log 25OHD level
Dudding T. 2015 (192)	caries experience	white	5545	0.93 (0.83, 1.05)	0.26	per 10 nmol/L increase 25OHD level
Dudding T. 2015 (192)	detanl GA	white	4072	0.96 (0.75, 1.22)	0.72	per 10 nmol/L increase 25OHD level
Dudding T. 2015 (192)	early caries onset	white	1933	1.09 (0.89, 1.34)	0.37	per 10 nmol/L increase 25OHD level
Li SS. 2016 (193)	Lumar 1-4 BMD	Chinese	1824	-0.048 (0.056)	0.384	g/cm2_per unit increase in log-transformed 25OHD
Li SS. 2016 (193)	Femoral neck BMD	Chinese	1824	-0.044(0.039)	0.261	g/cm2_per unit increase in log-transformed 25OHD
Li SS. 2016 (193)	Total hip BMD	Chinese	1824	-0.041 (0.042)	0.326	g/cm2_per unit increase in log-transformed 25OHD
Li SS. 2016 (193)	PTH	Chinese	1824	0.088 (0.062)	0.152	pg/mL_per unit increase in log-transformed 25OHD
Li SS. 2016 (193)	P1NP	Chinese	1824	-0.099 (0.098)	0.312	g/L_per unit increase in log-transformed 25OHD
Tyrrell J. 2016 (194)	birth weight	white	30340	-26 (-54, 2)	0.13	g per 10% lower 25OHD level
Vimaleswaran KS. 2013 (195)	BMI	white	123864	-0.002 (-0.009, 0.005)	0.57	per risk allele
Vimaleswaran KS. 2013 (195)	BMI	white	123864	0.002 (-0.006, 0.009)	0.67	per risk allele

Hysinger EB. 2016 (196)	pediatric asthma	white	5080/1203	-0.0000351	0.85	NA
Hysinger EB. 2016 (196)	severe asthma exacerbations	white	NA	-0.00833	0.86	NA
Liefaard MC. 2015 (197)	LnCRP	white	10788	-0.018	0.082	1 SD change in 25OHD level
Ooi EM. 2014 (198)	remnant cholesterol	white	79743	4.0 (-2.4, 11)	0.22	% per 50% decrease in 25OHD level
Ooi EM. 2014 (198)	LDL cholesterol	white	79812	2.2 (-1.7, 6.2)	0.28	% per 50% decrease in 25OHD level
Ooi EM. 2014 (198)	HDL cholesterol	white	85363	-6.0 (-10, -2.3)	0.001	% per 50% decrease in 25OHD level

Chapter II: Introduction – PheWAS

A systematic literature review of PheWAS studies

There have been many PheWAS studies prior to my study. Thus, it is necessary to summarize the study design, methods applied and main results of the pervious PheWAS studies in order to understand the rational and process of conducting a PheWAS study and to setup the PheWAS methodology that I will apply before original data analyses. In this chapter, I conducted a systematic literature review of all published PheWAS studies, aiming to explore common PheWAS designs, statistical methods, the workflow for conducting and reporting PheWAS, and the advantages, disadvantages and future development of this approach.

2.1 Methodology

2.1.1 Literature search strategy and inclusion and exclusion criteria

I searched Medline and Embase bibliographic databases using the terms "phenome", and "wide", and "association" and "stud*", or "PheWAS" up to end of 2017, which resulted in 305 papers. I did not include any reviews, correspondence, conference abstracts, literature introducing methodology or algorithm/software, literature which were not actual PheWAS, literature without full text, or research experiments conducted in animals or animal/ human cell lines. I also only retained papers written in English. With Xue Li, my PhD peer, we independently reviewed the literature and any discrepancies were solved by discussion.

2.1.2 Data extraction

From every study, I abstracted the following information: cohort name, sample size, ethnicity, age, type of data (i.e. EMR, clinical trial or large biobank), way of phenotyping (i.e. ICD curated/holistic, number of PheWAS groups, whether the rule of 2 was used, and the least number of cases), multiple comparison testing, regression covariates, method of association test (i.e. logistic regression, linear regression, chi-square test, Fisher's exact test, or others), the statistical software and packages they used, key findings, the statistically significant novel findings or the findings they gave emphasis to in their texts.

2.2 Review results

2.2.1 Included studies and characteristics

Forty-five papers were eligible for inclusion in my review through my literature search (**Figure 8**). The main characteristics and methodology applied of the included studies are presented in **Table 8**. Main findings of the studies are presented in **Table 9**.

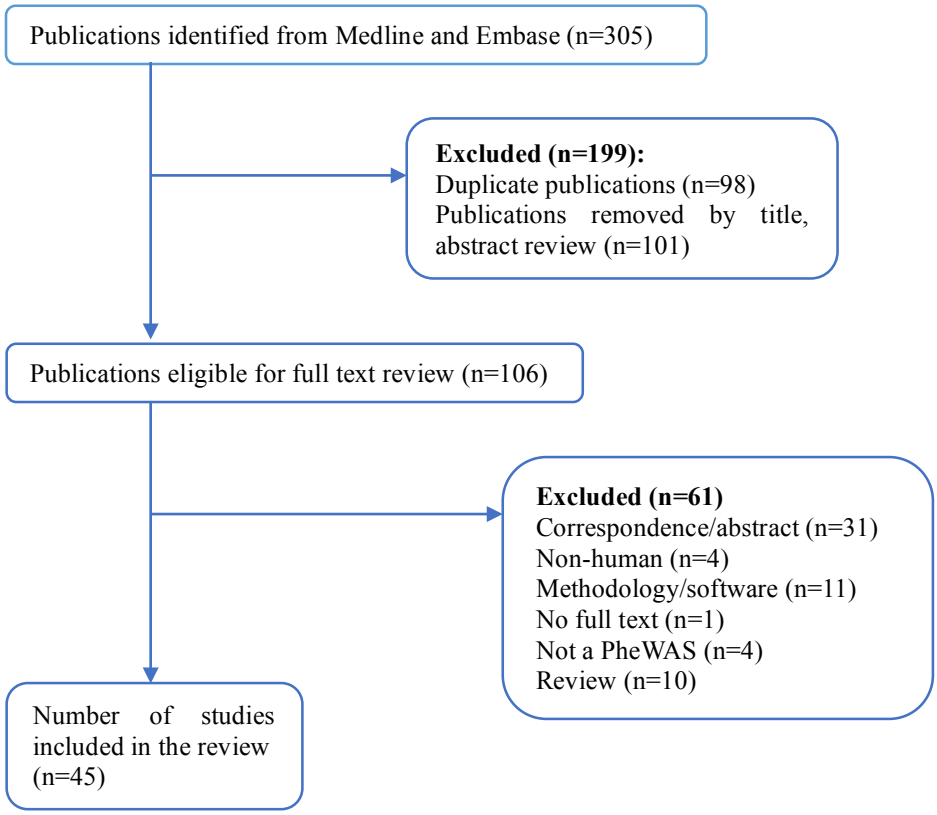


Figure 8. Flow chart of study selection process of the systematic literature review of PheWAS studies.

Table 8. Characteristics and method of included PheWAS studies.

Reference	Sample size	Study population	Predictors	Research Question and Design	Phenotyping method applied	Multiple testing method	Covariates included
PheWAS based on EMRs cohort							
Denny JC 2010 (199)	6,005	European American	rs3135388 rs17234657 rs2200733 rs1333049 rs6457620	Proof of concept of a high throughput phenotyping algorithm, 5 SNPs with known disease associations were genotyped	ICD curated [a] 25 as least number of cases [b] 776 phenotype groups	Bonferroni	NA
Denny JC 2011 (200)	13,617	European American	rs965513 (*FOXE1*)	A GWAS for primary hypothyroidism, followed by a PheWAS for significant locus, 9q22 near *FOXE1*	ICD curated rule of 2 [c] 20 as least number of cases 957 phenotype groups	Bonferroni	age sex first 3 PCs hypothyroidism
Pathak J 2012 (201)	6,307	Multiple (>90% white)	rs5219 rs7903146 rs12255372 rs13266634	To show the use of Resource Description Framework for representing EMR diagnoses and procedure data, and enable federated querying via standardized Web protocols to identify subjects	ICD curated 25 as least number of cases 516 phenotype groups	NA	NA

Denny JC 2013 (202)	13,835	European descent	3144 SNPs in NHGRI's GWAS catalogue	The first large scale PheWAS, exploring associations between 3144 SNPs and 1358 EMR-derived phenotype groups in a sample of 13835 participants	ICD curated rule of 2 25 as least number of cases 1358 phenotype groups	FDR	age gender study site first 3 PCs
Hebbring SJ 2013 (203)	4,235	white non-Hispanic	rs3135388 (*HLA-DRB1**1501)	To study the pleiotropic nature of *HLA-DRB1*, associations between rs3135388 and 4841 phenotype groups were tested.	ICD holistic [d] rule of 2 9 as least number of cases 4841 phenotype groups	Bonferroni	age sex year of EMR
Liao KP 2013 (204)	2,526	European ancestry	autoantibodies (ACPA, ANA, anti-TPO)	To study the association between autoantibodies and clinical diagnoses from EMR among Rheumatoid Arthritis cases and controls	ICD curated outcomes with prevalence >1% were analyzed	Bonferroni	age sex
Neuraz A 2013 (205)	442	NA	thiopurine S-methyltransferase (TPMT) activity	To test the association between clusters of TPMT activity and diagnoses from ICD10 codes	ICD curated 5 as least number of cases 156 groups (ICD10) or	FDR	NA

					289 groups (ICD9)		
Ritchie MD 2013 (206)	13,859	European American	23 SNPs in five loci associated with QRS duration	A GWAS for markers associated with QRS duration, and a subsequent PheWAS to search for associated diagnoses.	ICD curated rule of 2 (actual chose 4) 50 as least number of cases 778 phenotype groups	Bonferroni	age sex first 3 PCs
Carroll RJ 2014 (207)	6,005	European American	rs3135388 (*HLA-DRB*) white blood cell count	A proof of concept for a PheWAS R package	ICD curated 20 as least number of cases 1127 phenotype groups	NA	age sex
Cronin RM 2014 (208)	24,198	European	54 *FTO* SNPs previously associated with BMI and/or T2D	To study the pleiotropy of variants in *FTO* gene	ICD curated rule of 2 20 as least number of cases 1645 phenotype groups	FDR Bonferroni	age sex first 3 PCs study site

Mitchell SL 2014 (209)	11,519	African American	mtSNPs	To study the relationship between mtDNA variants and phenotypes	targeted: 8 cardiac related traits	NA	age sex first 2 PCs
Namjou B 2014 (210)	4,268	European Ancestry	2476 SNPs from previously published GWAS studies	A PheWAS in pediatric cohort for previously identified SNPs	ICD curated rule of 2 20 as least number of cases 539 phenotype groups	FDR permutation	first 2 PCs
Shameer K 2014 (211)	13,688	Asians Pacific Islanders	81 SNPs associated with PLT, MPV or both	A GWAS on MPV and PLT, followed by a PheWAS	ICD curated rule of 2 25 as least number of cases 1368 phenotype groups	Bonferroni	age sex
Boland MR 2015 (212)	1,749,400	Multiple	Birth month	Systematically exploring the relationships between birth month and risk for 1688 health conditions.	SNOMED-CT >1000 cases 1688 phenotypes	FDR	NA
Diogo D 2015 (213)	29,377	European ancestry	rs34536443 rs35018800 rs12720356	Fine map the 19p13 locus (RA risk locus) to find causal variants, and explore the pleiotropic effect by PheWAS	ICD curated rule of 2 outcomes with prevalence >1%	Bonferroni	age sex PCs

					were analyzed 502 phenotype groups		
Namjou B 2015 (214)	1067 children; 2227 adults	European	TA repeat in *UGT1A1*	GWAS of serum bilirubin levels, followed by PheWAS	ICD curated 377 outcomes	NA	Age, gender, genotyping platforms, PCs
Ye Z 2015 (215)	14,875	Native American	105 presumed functional (stop-gain and stop-loss) variants	PheWAS on functional variants	ICD holistic rule of 2 8 as least number of cases 4841 phenotype groups	NA	sex year of EMR
Hebbring SJ 2015 (216)	42,35	98% white or non-Hispanic	5 SNPs with previously known associations	PheWAS using clinical text data in phenotyping	clinical text based 50 as least number of cases	NA	NA
Li L 2016 (217)	1,169,599	Multiple	Birth month	Replicating the publication by Boland MR 2015	SNOMED-CT >1000 cases 1433 phenotypes	FDR	NA
Liu J 2016 (218)	7,481	Multiple	2692 human major	PheWAS for 2692 MHC variants	ICD holistic	Independent replication	Sex Year of EMR

			histocompatibility complex (MHC) variants		rule of 2 9 as least number of cases 6221 phenotype groups		
Millwood IY 2016 (219)	91,428	Chinese	rs76863441	PheWAS for rs76863441, associated with lipoprotein-associated phospholipase A_2	ICD-coded, following the WHO ICD10 structure	Bonferroni	Sex, region, age
Mosley JD 2016 (220)	29,349	European	SNPs in the human leukocyte region	PheWAS using linear mixed model	ICD curated rule of 2 50 as least number of cases	FDR	Age, sex, 20 PCs
Oetjens MT 2016 (221)	6,892	European and African Americans	184 functional variants in 34 pharmacogenes	Identify the pleiotropy of pharmacogenes.	ICD curated rule of 2 35 as least number of cases	Bonferroni	Age sex
Verma A 2016 (222)	5,923	European	76,861 SNPs	PheWAS for variants in immune related loci	ICD holistic rule of 3 10 as least number of cases	Bonferroni	Age sex first 5 PCs

Verma A 2016 (223)	12,039	European	286 SNPs	GWAS for 21 lab measurement extracted from EMR, followed by PheWAS for significant SNPs	ICD holistic rule of 3 200 as least number of cases 165 phenotype groups	Bonferroni	Sex age age-squared BMI first 4 PCs
Verma A 2016 (224)	41,057	Multiple	25 stop-gain variants	PheWAS for stop-gain variants	ICD holistic rule of 3 10 as least number of cases	NA	Sex study site genotyping platform first 3 PCs
Verma A 2016 (225)	45,899	European	687 SNPs	GWAS for 25 laboratory traits, followed by PheWAS stratified by variance of laboratory traits (high variance and low variance individuals were analysed separately)	ICD holistic rule of 3 200 as least number of cases 541 phenotype groups	NA	Sex age First 4 PCs
Cortes A 2017 (226)	152,732	Multiple	*HLA-B*27:05* allele	PheWAS for *HLA* variants using a Bayesian analysis framework	ICD holistic	NA	NA
Doss J 2017 (227)	2,199	Multiple	Seropositive RA vs seronegative RA;	PheWAS for rheumatoid arthritis subgroups	ICD curated rule of 2 20 as least number of cases	FDR	Sex age

			RF-positive vs RF-negative; ACPA-positive vs RA-negative				
Jannot AS 2017 (228)	126,736	NA	Hospital acquired acute kidney injury (HA-AKI)	PheWAS for HA-AKI	ICD holistic 1513 ICD-10 groups	Bonferroni	NA
Karnes JH 2017 (229)	37,270	European	1,164 *HLA* variants	PheWAS for *HLA* variants	ICD curated rule of 2 40 as least number of cases 1368 phenotypes	FDR	Sex age first 2 PCs
Liao KP 2017 (230)	1,006	Multiple	36 autoantibodies	PheWAS for associations between autoantibodies implicated in RA and clinical phenotypes	ICD curated outcomes with prevalence >3% 206 phenotype groups	FDR	Sex age race
Liu J 2017 (231)	14,275	Multiple	*SULT1A1* copy number variation	ICD-9 code and text-based PheWAS for *SULT1A1* copy number variation	ICD holistic 9 as least number of cases 6910 phenotypes;	Bonferroni	Sex length of EMR

Robinson R 2017 (232)	2,907	Multiple	loxoscelism	PheWAS for phenotypic associations with loxoscelism.	text-based phenotyping 23,382 text terms ICD curated rule of 2 20 as least number of cases	Bonferroni	Sex age
PheWAS based on other cohorts							
Pendergrass SA 2013 (233)	70,061	Multiple	83 GWAS-derived variants	A comprehensive PheWAS	NA	inter-database replication	ethnicity
Hall MA 2014 (234)	14,042	Multiple	80 SNPs	Comprehensive test of associations between 80 SNPs and 1008 phenotypes, stratified by ethnicity	NA	inter-database replication	ethnicity
Millard LA 2015 (235)	8,121	European children	BMI allele score including 32 loci	MR-PheWAS: exploring causal associations between BMI and broad range of outcomes	172 continuous variables	NA	NA

Moore CB 2015 (236)	2,547	Black White Hispanic	5,954,294 SNPs	To identify pharmacogenomic associations, a PheWAS in a HIV trial data, recruiting pretreatment individuals only.	27 pretreatment and baseline laboratory assays	inter-database replication	age sex first 5 PCs CD-4 cell counts (square root transformed)
Roesch SL 2015 (237)	128	Mexican American Hispanics	Fibroblast Growth Factors 19 and 21	PheWAS between clusters of FGF19 and FGF 21 levels and 205 clinical variables	205 clinical variables	Bonferroni	NA
Karaca S 2016 (238)	974	Turkish children	*ADAM33, ADRB2, CD14, IL13, IL4, IL4R, MS4A2, SERPINE1, TNF*	PheWAS for 9 immune related genes	116 allergy relevant traits	Locus-specific Bonferroni	Age, sex, sampling centre
Krapohl E 2016 (239)	3152	UK adolescents	13 scores for cognition and psychiatric disorders	PheWAS for 13 genome-wide polygenic scores	50 cognitive traits or psychiatric diseases	NA	NA
Polimanti R 2016 (240)	26,394	Multiple	CHRNA3-*CHRNA5 locus,*	Identify novel traits associated with alcohol or nicotine use related variants.	360 traits	Bonferroni	age age squared first 10 PCs

			ADH1B, ALDH2				
Ehm MG 2017 (241)	521,000	NA	7 variants in Th17 and IL-17 pathway	A PheWAS with self-reported data for Th17 and IL-17 pathway.	1254 self-reported phenotypes	FDR	Sex age first 5 PCs
Polimanti R 2017 (242)	11,271	Multiple	rs113288603 (*CYP2A6*)	PheWAS for *CYP2A6* locus	358 traits	FDR	Sex age age-squared first 10 PCs
Verma A 2017 (243)	1181	AIDS patients Multiple ethnicities	2,544 SNPs	Testing associations for laboratory phenotypes among antiretroviral treatment naïve patients	774 phenotypes	Permutation	Sex age first 10 PCs

a. ICD curated, ICD diagnosis codes were binned into phenotype categories due to predefined criteria.

b. the required least number of cases for a phenotype to be included in statistical analysis.

c. to be defined as case for a specific phenotype group, a participant must have at least two occurrence of related ICD codes on separate episodes.

d. ICD holistic, the phenotype group in statistical analysis is simply defined by ICD codes, without any binning.

Abbreviations: EMR, electronic medical record; ICD, International Statistical Classification of Diseases and Related Health Problem; least number of cases, phenotype groups with case numbers under which will not be analyzed; PC, ancestral principle components based on genotypic information, used to adjust for population stratification; FDR, false discovery rate; SNP, single nucleotide polymorphism; mtSNP, mitochondrial single nucleotide polymorphism.

Table 9. Main findings of included PheWAS studies

Reference	Predictors	Key findings
PheWAS based on EMRs cohort		
Denny JC 2010 (199)	rs3135388 rs17234657 rs2200733 rs1333049 rs6457620	• Four of seven known associations were replicated. • 19 novel associations identified.
Denny JC 2011 (200)	rs965513 (*FOXE1*)	• The strongest association was with hypothyroidism. • Additional associations included: thyroiditis, nodular, multinodular goiters, and thyrotoxicosis. • Associations with Graves' disease and thyroid cancer were not significant.
Pathak J 2012 (201)	rs5219 rs7903146 rs12255372 rs13266634	• All of four SNPs were associated with diabetes and related traits. • Did not replicate associations between rs12255372 and breast cancer or prostate cancer. • All of four were associated with skin and tissue related diseases.
Denny JC 2013 (202)	3144 SNPs in NHGRI's GWAS catalogue	• Successfully replicated 66% of previous identified associations. • 63 novel associations were identified surviving multiple testing correction, the strongest of which was actinic keratosis and rs12203592 (*IRF4*).

Hebbring SJ 2013 (203)	rs3135388 (*HLA-DRB1**1501)	• None of the results met a conservative Bonferroni correction threshold. • The established association with MS was replicated. • The strongest association was with alcohol cirrhosis of the liver. • Also identified associations with erythematous conditions, benign neoplasms of the respiratory and intrathoracic organs, and benign neoplasm of other parts of the digestive system (replicating previous findings).
Liao KP 2013 (204)	autoantibodies (ACPA, ANA, anti-TPO)	• anti-TPO was associated with hypothyroidism in both cases and controls; and associated with thyroiditis in controls. • The presence of ANAs was significantly associated with a diagnosis of Sjogren's/Sicca syndrome in RA cases; and other chronic non-alcoholic liver disease in controls.
Neuraz A 2013 (205)	thiopurine S-methyltransferase (TPMT) activity	• very high TPMT activity was found to be associated with iron-deficiency anemia and diabetes mellitus.
Ritchie MD 2013 (206)	23 SNPs in five loci associated with QRS duration	• None of the results survived multiple testing. • The most significant associations were between rs6795970 (*SCN10A*) and cardiac arrhythmias or atrial fibrillation and flutter, which were independent of QRS duration.
Carroll RJ 2014 (207)	rs3135388 (*HLA-DRB*) white blood cell count	• rs3135388 was associated with multiple sclerosis. • white blood cell count was associated with infections, leukemias, myeloproliferative diseases and anemia.

Cronin RM 2014 (208)	54 *FTO* SNPs previously associated with BMI and/or T2D	• rs8050136 was found to be associated with obesity, morbid obesity, T2D, OSA, NAFLD and fibrocystic breast disease. After adjustment for BMI, associations with obesity, T2D and OSA were highly attenuated. • Associations between rs16952520 and non-inflammatory disorders of cervix, rs7199182 and chronic periodontitis, were not affected by adjustment, and they were also not associated with obesity or T2D.
Mitchell SL 2014 (209)	mtSNPs	• 7 mtSNPs were found to be associated with total cholesterol. • 13 mtSNPs were found to be associated with T2D. • More SNPs were significantly associated than would be expected by chance alone for total cholesterol and T2D. • mt16189 was found to be associated with T2D, which was previously reported only in Asian and European descents.
Namjou B 2014 (210)	2476 SNPs from previously published GWAS studies	• Many previous associations were replicated, including JRA, thyroiditis and T1D. • Several novel findings were identified, the strongest of which were between *PLCL1* (best SNP 1595825) and developmental delays and speech disorder, *IL5-IL13* region (best SNP rs12653750) with Eosinophilic Esophagitis.
Shameer K 2014 (211)	81 SNPs associated with PLT, MPV or both	• The strongest associates were between rs3819299 and inflammatory spondylopathies, ankylosing spondylitis and non-infectious uveitis.
Boland MR 2015 (212)	Birth month	• 55 diseases statistically significantly associated with birth month. • 19 diseases were previously reported by literatures, 20 diseases were closely related to the previously reported ones, 16 birth month-disease associations were novel.

Diogo D 2015 (213)	rs34536443 rs35018800 rs12720356	• Three SNPs were found to be protective against RA, SLE and suggestive IBD. • No convincing evidence for association with complex phenotypes other than autoimmune diseases were found.
Namjou B 2015 (214)	TA repeat in *UGT1A1*	• None of the associations were statistically significant controlling for multiple corrections. • There were toward trend effects on cerebrovascular disease and ischemic stroke in adults.
Ye Z 2015 (215)	105 presumed functional (stop-gain and stop-loss) variants	• The most significant association reported was between rs3731608 and current long-term drug use, but it was not replicated in another independent set. • A nonsense variant rs2736911 was found to be associated with age-related macular degeneration.
Hebbring SJ 2015 (216)	5 SNPs with previously known associations	• The association between rs1061170 and age-related macular degeneration (specifically, with words strings "macular degeneration", "non-exudative" and "exudative"), as well as word string "visudyne", a drug commonly used to treat age-related macular degeneration, survived multiple testing.
Li L 2016 (217)	Birth month	• Four circulatory outcomes were significant at FDR adjusted level, including coronary arteriosclerosis, essential hypertension, angina and pre-infarction syndrome.

Liu J 2016 (218)	2692 human major histocompatibility complex (MHC) variants	• Eight novel outcomes were found to be associated with MHC region. These include unspecific histoplasmosis retinitis, haemangioma of intra-abdominal structures, pneumonia due to staphylococcus, lichen planus, dyshidrosis, other and unspecified nonspecific immunological findings, infraspinatus (muscle) (tendon) sprain, contusion of wrist.
Millwood IY 2016 (219)	rs76863441	• None of the outcomes were statistically significantly associated with rs76863441.
Mosley JD 2016 (220)	SNPs in the human leukocyte region	• 44 phenotypes were associated with HLA variants.
Oetjens MT 2016 (221)	184 functional variants in 34 pharmacogenes	• Previously known association between rs2231142 and gout, and association between rs4149056 and jaundice were replicated. • Novel associations between rs1143672 and renal osteodystrophy were identified.
Verma A 2016 (222)	76,861 SNPs	• rs6910071 (C6orf10) were significantly associated with rheumatoid arthritis. • rs2239167 (ATN1) were significantly associated with type 2 diabetes.
Verma A 2016 (223)	286 SNPs	• 39 associated survived the Bonferroni correction. The most significant one was between rs9273363 (in HLA region) and type 1 diabetes.

		• The majority of the PheWAS phenotype groups highly related to the clinical lab measures associated with same SNPs.
Verma A 2016 (224)	25 stop-gain variants	• Previously known associations were replicated. • rs328 (*LPL*) was found to be associated with disorder of lipoid metabolism. • rs1137617 (*KCNH2*) was found to be associated with acquired hypothyroidism. • rs12060879 (*DPT*) was found to be associated with complications of peculiar to certain specified procedures.
Verma A 2016 (225)	687 SNPs	• 717 PheWAS associations were found at *P* level of 0.001. • 39 SNPs associated with type 1 diabetes in patients with high variance of plasma glucose levels, but not in patients with low variance. • 4 SNPs in *UMOD* were associated with chronic kidney disease in patients with high variance for aspartate aminotransferase.
Corte A 2017 (226)	*HLA-B*27:05* allele	• Their Bayesian analysis framework increased statistical power by more than 20%. • Bayesian analysis identified associations not found by traditional PheWAS methods, including association between *HLA-B*27:05* and 145 ICD-10 terms.
Doss J 2017 (227)	Seropositive RA vs seronegative RA; RF-positive vs RF-negative; ACPA-positive vs RA-negative	• Seronegative RA was associated with myalgia and myositis, fibromyalgia, and back pain. • Seropositive RA was associated with chronic airway obstruction and tobacco use.

Jannot AS 2017 (228)	Hospital acquired acute kidney injury (HA-AKI)	• Surgical procedures and hemodynamic impairment were found to be the main risk factors for HA-AKI.
Karnes JH 2017 (229)	1,164 *HLA* variants	• 1,955 significant allele-phenotype associations were found for HLA variants. • *HLA-DQB:1*03:02* was associated with type 1 diabetes. • *HLA-B*27* was associated with ankylosing spondylitis.
Liao KP 2017 (230)	36 autoantibodies	• 24 significant associations were identified. • Autoantibodies against fibronectin was associated with obesity. • Autoantibodies against fibrinogen was associated with pneumonopathy.
Liu J 2017 (231)	*SULT1A1* copy number variation	• No phenotype passed the Bonferroni correction from the ICD-9 based PheWAS. • In the text-based PheWAS, association with term 'Nasacort' passed Bonferroni threshold.
Robinson 2017 (232)	loxoscelism	• 29 associations were significant, including rash, toxic effect of venom, and haemolytic anemia.
PheWAS based on other cohorts		

Pendergrass SA 2013 (233)	83 GWAS-derived variants	• 111 associations were significant for the same ethnicity, SNP and phenotype-class across two or more study sites. • 52 replicated associations, 26 represented phenotypes closely related to previous associations, 33 potentially novel associations. • Most significant novel finding: association between rs1333049 (*CDKN2A/B*) and haemoglobin levels in African Americans, which was found to be associated with type 2 diabetes in European Americans previously.
Hall MA 2014 (234)	80 SNPs	• 39 replicated associations. • 9 related to reported associations. • 21 novel associations. • 13 SNPs showed evidences of pleiotropy.
Millard LA 2015 (235)	BMI allele score including 32 loci	• A total of 21 outcomes were associated with BMI at $P < 0.05$ level, after Bonferroni correction, only association between BMI and HDL at age 9 kept significant. • A novel effect of BMI on global self-worth score were suggested.
Moore CB 2015 (236)	5,954,294 SNPs	• 20 SNP-phenotype pairs matched identical or very close related GWAS catalog associations. • 23 SNPs with 29 associations which differed considerably from GWAS catalog, including rs10494326 with neutrophil count and rs2201841 with plasma chloride concentrations.

Roesch SL 2015 (237)	Fibroblast Growth Factors 19 and 21	• 21 variables associated with either FGF19 or FGF21, never both. Always, higher FGF21 or lower FGF19 were associated with some variables. After Bonferroni, only high glucose and high FGF21 kept significant.
Karaca S 2016 (238)	21 SNPs in 9 immune related loci	• rs2280090 (*ADAM33*) was associated with MEF240%, allergic bronchitis. • rs3918396 (*ADAM33*) was associated with risk of wheezing and eczema comorbidity. • rs2243250 (*IL4*) was associated with FEV240. • rs2569190 (*CD14*) was associated with diagnosis of asthma.
Krapohl E 2016 (239)	13 scores for cognition and psychiatric disorders	NA
Polimanti R 2016 (240)	*CHRNA3-CHRNA5 locus, ADH1B, ALDH2*	• Prior association between *ADH1B* and drinking behaviour was replicated. • Novel associations between *ADH1B* and psychological traits, socioeconomic status, vascular/metabolic conditions, and reproductive health were suggested. • Prior associations between *CHRNA3-CHRNA5* and smoking status, lung cancer and asthma were replicated. • Novel associations between *CHRNA3-CHRNA5* and high-cholesterol-median use, distrustful attitude were suggestive.
Ehm MG 2017 (241)	7 variants in Th17 and IL-17 pathway	• *TYK2* was associated with tonsillectomy, strep throat occurrences and teen acne. • *IL23R* was associated with dandruff frequency. • *TRAF3IP2* was associated with risk of male-pattern balding. • *RORC* (variant rs4845604) was associated with protection from allergries.

Polimanti R 2017 (242)	rs113288603 (*CYP2A6*)	• rs113288603 was associated with hearing loss symptoms in nicotine-exposed elderly subjects significantly. • In non-nicotine-exposed elderly subjects, this association was not significant.
Verma 2017 (243)	2,544 SNPs	• rs12683493 was associated with atazanavir pharmacokinetics. • rs2368393 was associated with CD4 T-cell count. • rs7865618 was associated with HIV-1 RNA phenotypes. • Both previously reported and possibly novel associated were identified.

Abbreviations: *IRF4*, interferon regulator factor 4; MS, multiple sclerosis; anti-APO, anti-thyroidperoxidase antibody; ANAs, antinuclear antibodies; RA, rheumatoid arthritis; TPMT, thiopurine S-methyltransferase; *SCN10A*, sodium voltage-gated channel alpha subunit 10; OSA, obstructive sleep apnoea; NAFLD, non-alcoholic fatty liver disease; JRA, juvenile idiopathic arthritis; *PLCL1*, phospholipase C like 1; *IL5*, interleukin 5; *IL13*, interleukin 13; PLT, number of circulating platelets; MPV, mean platelet volume; SLE, systemic lupus erythematosus; IBD, inflammatory bowel disease; *CDKN2A/B*, cyclin-dependent kinase inhibitor 2A/B; FGF19, fibroblast growth factor 19; FGF21, fibroblast growth factor 21; T1D, type 1 diabetes.

In 2010, Denny et al. conducted a state-of-the-art PheWAS in a sample of 6,005 European Americans. In this study, they first proposed the phecode system. This is an automatic code translation table working with ICD9 billing codes. Related ICD9 codes representing the same disease were classified into the same identical phecode group by the algorithm. All participants were linked to their EMR data, and a total of 776 different disease groups were defined by the phecode system. In their study, five genetic polymorphisms were genotyped in order to validate the ability of PheWAS to replicate established associations and find novel associations. The genetic predictors included rs1333049 (associated with coronary artery disease and carotid artery stenosis), rs2200733 (associated with atrial fibrillation), rs3135388 (associated with multiple sclerosis and systemic lupus erythematosus), rs6457620 (associated with rheumatoid arthritis) and rs17234657 (associated with Crohn's disease). For every distinct phenotype group, they calculated case and control genotype distributions and the χ^2 distribution, associated *P* values and allelic OR values. Rarer phenotypes, whose case numbers were less than 25, were not included into the analyses. Out of the seven previously established associations, four were replicated by their PheWAS (rs3135388 with multiple sclerosis, rs17234657 with Crohn's disease, rs1333049 with coronary artery disease and rs6457610 with rheumatic arthritis). In addition, 19 novel associations were identified for these five variants at a threshold of $P < 0.001$. However, none of them survived the Bonferroni correction (199). The results of this study supported the capacity of PheWAS for replicating known associations and further providing statistical evidences on possibly novel associations.

Following their first PheWAS study, Denny et al. published another one in 2011. In this study, they first conducted a GWAS for hypothyroidism in 1317 cases and 5053 controls. Four SNPs (rs7850258, rs965513, rs925489, and rs10759944) near gene *FOXE1* were found to be statistically significantly associated with hypothyroidism. In a subsequent PheWAS, rs965513 was selected and its association with 957 phenotypes defined by the phecode system were tested comprehensively in 13,617 individuals. In this PheWAS analysis, they required each case to have at least two ICD codes in a PheWAS case group, which could act to increase positive predictive value. Phenotypes with less than 20 cases were excluded. Logistic regressions were ran adjusting for age,

sex, the first three ancestral principle components and hypothyroidism status. Hypothyroidism (OR = 0.76, $P = 2.7*10^{-13}$), thyroiditis (OR = 0.58, $P = 1.4*10^{-5}$), nodular goitres (OR = 0.76, $P = 3.1*10^{-5}$), multinodular goitres (OR = 0.69, $P = 3.9*10^{-5}$), thyrotoxicosis (OR = 0.76, $P = 1.5*10^{-3}$) were associated with rs9655133. However, Graves' disease (OR = 1.03, $P = 0.82$) and thyroid cancer (OR = 1.29, $P = 0.09$) were not associated with the SNP (200).

In another PheWAS, data were stored and represented within the Resource Description Framework (RDF). RDF provides a powerful framework for expressing and integrating any type of data by representing data as labelled graphs. Therefore, it allows fast querying and information retrieval across multiple sources of data, including genotype data, EMR data and lab data. This study selected four SNPs to be associated with T2D, including rs5219, rs7903146, rs12255372 and rs13266634, and studied their associations with all disease and procedure types in 6,307 T2D cases. In this study, researchers first extracted all ICD9 billing codes, and defined phenotype groups with Clinical Classification Software (CCS) (244) based on ICD9 codes. CCS classified over 14,000 diagnoses codes and 3,900 procedure codes into 285 and 231 distinct diagnosis and procedure phenotype groups, respectively. Similarly, phenotypes with less than 25 cases were not analysed in their PheWAS. From their analyses, all 4 SNPs were associated with diabetes and related traits. In addition, all 4 SNPs were associated with skin and tissue related diseases (e.g., corns, seborrheic dermatitis), which had not been reported by previous studies. At last, although the association between rs12255372 and breast cancer (245) and prostate cancer (246) had been found by previous studies, this study failed replicating these associations, which might be caused by the relative small sample size (201).

In addition, Denny et al. conducted a very large scale PheWAS, which explored phenome wide associations for 3144 SNPs with known associations on the GWAS catalogue in 13,835 individuals, and published their findings in year 2013 on Nature Biotechnology. EMR data were retrieved and the coded into 1,358 unique phenotypes by the phecode system. Two instances of related ICD billing codes were required for an individual to be coded as a case for specific phenotype group; Groups with less than

25 cases were not analysed. Their PheWAS successfully replicated 66% of previously reported genotype-phenotype associations (51/77) at *P* level of 0.05. Sixty-three novel statistically significant associations (surviving false discovery rate correction) were identified. The strongest novel association observed was between rs12203592 (*IRF4*, which was previously reported to be associated with hair and eye colour) and actinic keratosis (OR = 1.69, P = $4.1*10^{-26}$) (202).

Hebbring et al. studied the associations between the rs3135388 polymorphism and 4,841 phenotypes in a sample of 4,235 individuals. In phenotyping, they did not classify ICD billing codes into phenotype groups in the same way as the above studies. Alternatively, they used ICD9 codes directly as phenotypes. Each code represents a unique phenotype group, and individuals without records of the specific code were treated as controls for the code group. To be defined as a case for a coding group, an individual must have at least two occurrences of that code. Rare diseases whose case number were less that 9 were not analysed. In their study, none of the phenotypes survived a conservative Bonferroni correction. However, the previously established association with multiple sclerosis was suggested (P = 0.023) at P level of 0.05. Moreover, rs3135388 was found to be nominally associated with alcohol cirrhosis of the liver (P = 0.00011), erythematous conditions (P = 0.0054), benign neoplasms of the respiratory and intrathoracic organs (P = 0.042), and benign neoplasms of other parts of the digestive system (P = 0.0023) (203). Considering their limited sample size and the large number of phenotypes tested, the multiple testing burden was considerably large and it was hard for them to find any statistically significant association which survived multiple testing.

All of the above 5 PheWAS studies employing genetic variants as their predictors of interest. In addition to genetic factors, other exposures can also be used as predictors in PheWAS, such as plasma biomarkers. In another PheWAS, the phenome wide association for the presence of anti-citrullinated protein antibodies (ACPAs), antinuclear antibodies (ANAs), and anti-thyroid peroxidase (anti-TOP) antibodies were studied in a sample of 1,290 RA cases and 1,236 controls. The phecode system was used in defining phenotype groups, and only phenotypes with greater than 1%

prevalence were analysed. For RA cases, 512 phenotype groups were analysed and for controls, 698 groups were analysed. Logistic regression models were fitted adjusting for age and sex. From their results, the presence of anti-TPO antibodies was associated with hypothyroidism in both RA cases ($P = 1.2*10^{-16}$) and controls ($P = 9.2*10^{-10}$). Anti-TPO was also associated with thyroiditis in controls ($P = 1.2*10^{-7}$). The presence of ANAs antibodies was significantly associated with diagnosis of Sjogren's/Sicca syndrome in cases ($P = 8.6*10^{-6}$) and associated with other chronic non-alcoholic lever disease in controls ($P = 2.9*10^{-5}$) (204).

Another PheWAS with biomarkers as predictors was conducted by Neuraz et al. In their study, thiopurine S-methyltransferase activity (TPMTa) was classified into low TPMTa, normal TPMPa and very high TPMTa and association between categories of TPMTa and phenotypes were tested comprehensively in 442 individuals. Phenotyping was conducted in two different ways. Since the EMR of their individuals were in ICD10 codes, they first defined phenotype groups just based on first three digits of ICD10 billing codes (i.e. ICD10 holistic), and this led to 156 groups. In addition, they translated ICD10 codes back to ICD9 codes with the United Medical Language System, and then applied the phecode system, which generated 289 phenotype groups. Only phenotypes with no less than 5 cases were analysed. In the very high TPMTa versus other TPMTa (normal plus low TPMTa) group analysis, very high TPMTa was found to be associated with iron deficiency anaemia ($P = 0.0005$) and diabetes mellitus ($P = 0.0009$) from the ICD10 phenotype analysis. The results of ICD9 phenotype analysis were consistent with the ICD10 results (205).

In 2013, Ritchie et al. conducted another PheWAS following a GWAS. They first conducted a GWAS for QRS duration, which represents activation time in the cardiac ventricle and has been associated with adverse cardiovascular outcomes, in a sample of 5,272 European Americans. Twenty-three SNPs in 5 loci, which had been reported by previous GWAS, were replicated successfully by their GWAS. Then, they conducted a PheWAS for these 23 SNPs in 13,859 European Americans. They used the phecode system for their phenotyping, which generated 778 genotype groups. To be defined as a case for a specific phenotype group, an individual need to have at least

4 occurrences of related ICD9 codes. Phenotype groups of less than 50 cases were excluded from statistical analysis. Age, sex, and first 3 ancestral principal components were adjusted as covariates. In their study, none of the genotype-phenotype associations survived Bonferroni correction. The most significant associations were between rs6795970 (*SCN10A*) and cardiac arrhythmias ($P = 7.21*10^{-4}$) or arterial fibrillation and flutter ($P = 8.45*10^{-4}$), which were independent of QRS duration (206).

Denny et al. then produced a R package, which runs their phecode system, and published another PheWAS using the package. In this PheWAS, both genetic predictor (rs3135388 in *HLA-DRB*) and a non-genetic predictor (white blood count) were selected. In their study, 1,127 phenotype groups were generated by the phecode system. Phenotype groups with less than 20 cases were excluded from statistical analyses. For covariates, age and gender were adjusted. There had been another PheWAS study for rs3135388 on the same cohort (199) and the results from this study replicated known associations with multiple sclerosis found by the previous one (OR = 2.56, $P = 1.4*10^{-4}$ in this study; OR = 2.24, $P = 2.8*10^{-6}$ from previous PheWAS study), which demonstrated the robustness of this R package. In the PheWAS for white blood cell count, it was found to be associated with infections, leukaemia and other expected conditions (207).

In order to study the pleiotropy of genetic variants near gene *FTO*, which had been found to be associated with obesity and type 2 diabetes, Cronin et al. conducted a PheWAS in two populations of European ancestry. The first population comprised 10,487 individuals of European ancestry, and the other one comprised 13,711 individuals of European ancestry. PheWAS analyses were conducted in two populations separately, then a combined meta-analysis was conducted. PheWAS analyses were ran for 54 *FTO* SNPs. The previously described R package for phecode system was used in phenotyping. Cases for a given disease were defined as having at least two relevant ICD9 codes on different days. Analyses were conducted for only phenotypes occurring in at least 20 individuals. Analyses were first ran adjusting for age, sex, study site and first three ancestral principal components. Then, BMI was further adjusted to explore the association between variants and phenotypes

independent of BMI. The established associations between *FTO* (rs8050136) and obesity (OR = 1.25, 95% CI: 1.16-1.35, P = $2.10*10^{-9}$), and T2D (OR = 1.14, 95% CI: 1.08-1.21, P = $2.34*10^{-6}$) were successfully replicated by their study. Polymorphism rs8050136 was also significantly associated with sleep apnoea, but this was greatly attenuated after adjusting for BMI. Phenotypes which were associated with rs8050136 and this was not changed by adjustment of BMI included fibrocystic breast disease (OR = 0.84, 95% CI: 0.75-0.92, $P = 4.8*10^{-4}$) and non-alcoholic liver disease (OR = 1.19, 95% CI: 1.07-1.33, $P = 1.9*10^{-3}$). In addition, associations between rs16952520 and non-inflammatory disorders of cervix (OR=6.66, 95% CI: 3.03-14.64, $P = 2.36*10^{-6}$), associations between rs7199182 and chronic periodontitis (OR = 14.58, 95% CI: 3.97-53.57, $P = 5.40*10^{-5}$) were independent of BMI as well. Polymorphisms rs16952520 and rs7199182 were not associated with obesity or type 2 diabetes (208).

In another PheWAS, Mitchell et al. used variants on mitochondrial DNA (mtDNA) as their predictors. They studied the associations between 86 mtSNPs and 8 cardiovascular related traits, including BMI, total cholesterol, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides, mean corpuscular haemoglobin (MCH), type 2 diabetes and hypertension in 11,519 African Americans. Age, sex and the first two ancestral principal components were adjusted as covariates. Seven mtSNPs and 13 mtSNPs were found to be associated with total cholesterol and type 2 diabetes, respectively. More SNPs were statistically significantly associated than would be expected by chance alone for total cholesterol and type 2 diabetes. In addition, mt16189 was found to be associated with type 2 diabetes, which was previously reported only in Asians and European descents (209).

Namjou et al. conducted the first large-scale PheWAS in paediatrics. In their study, associations between 2,476 SNPs derived from published GWAS and 539 phenotypes were tested in 4,268 individuals of European ancestry. The phecode system was employed in phenotyping. An individual had to have two records of related ICD9 codes in order to be defined as case for specific phenotype group. Phenotype groups of less than 20 cases were not analysed. Logistic regression models were fitted, adjusted for the first two ancestral principal components. Known associations were

replicated, including associations between rs2476601 and juvenile rheumatoid arthritis (JRA), thyroiditis and type 1 diabetes, association between rs3806932 and Eosinophilic Esophagitis (EoE). Novel findings included associations between *PLCL1* (rs1595825) and developmental delays and speech disorders (OR = 0.65, 95% CI: 0.57-9.76, *P* = 1.13*10-8) and association between *IL5-IL13* (rs12653750) and EoE (OR = 1.73, 95% CI: 1.44-2.07, P = $3.03*10^{-9}$) (210).

In 2013, Shammer et al. conducted a GWAS which aimed to identify genetic variants affecting number of circulating platelets (PLT) and mean platelet volume (MPV) and explored the pleiotropic effect of significant SNPs by PheWAS. They had 13,582 individuals with data o PLT and 6,291 individuals with data on MPV. In their GWAS, they identified five chromosomal regions associated with PLT and eight regions associated with MPV. Eighty-one SNPs (56 associated with PLT, 29 associated with MPV and 4 SNPs associated with both) were studied in their subsequent PheWAS. They generated 1,368 phenotypes with the phecode system. At least two instances of related ICD9 billing codes were required for definition of cases. Rare phenotypes with less than 25 cases were not analysed. Age and gender were adjusted as covariates. Multiple autoimmune and haematological conditions were associated with these SNPs. The strongest associations were between rs3819299 and ankylosing spondylitis (*P* = $3.3*10^{-7}$), inflammatory spondylopathies (*P* = $5.7*10^{-8}$) and non-infectious uveitis (*P* = $4.6*10^{-7}$) (211).

In 2015, Bolland et al. conducted a PheWAS for birth month, which was suggested to impact health outcomes of multiple categories, among 1,749,400 individuals. They derived ICD-9 diagnosis codes of participants from their EMRs, and mapped ICD9 codes to Systemized Nomenclature for Medical-Clinical Terms (SNOMED-CT) codes. SNOMED-CT codes capture more clinical content compared with ICD-9 codes, thus is an alternative curating method for the 'phecode' system. In their PheWAS, a total of 1688 diseases were tested against their association with birth month. Fifty-five diseases were statistically significantly associated with birth month, among which 16 were novel associations which had not been reported by previous studies (212).

In 2015, Diogo et al. first fine mapped the RA risk locus at 19p13 to define causal variants, and then studied the pleiotropy of theses variants with PheWAS. In their study, three protein coding variants (P1104A, A928V and I684S) in gene *TYK2* were found to be associated with RA susceptibility. PheWAS was conducted in a sample of 29,377 white individuals. The phecode system was employed in phenotyping. A person needed to have two occurrences of related ICD9 codes in order to be defined as a case of a specific phenotype group. Outcomes with prevalence of more than 1% were analysed (502 phenotype groups), and logistic regression models were fitted adjusting for age, sex and ancestral principal components. The only statistically significant association from their PheWAS was between P1104A and RA (OR = 0.65, $P = 2.3*10^{-5}$). Thus, their study did not support any association between RA related variants and phenotypes other than RA (213).

In 2015, Namjou and colleagues conducted a GWAS for total serum bilirubin level and other liver function tests and found that a TA repeat in *UGT1A* gene was significantly associated with total serum bilirubin level. Then they conducted a PheWAS for this TA repeat in 1067 children and 2227 adults of European ancestry. A total of 377 phenotypes generated by the 'phecode' systems were tested. No phenotype survived multiple testing correction. There were toward trend effects on cerebrovascular disease and ischemic stroke in adults (214).

Most previous PheWASs studied SNPs identified by GWAS studies. However, top SNPs from GWAS were just tagging variants, and they did not necessarily have any biological functions. Causal variants, although they may have larger *P* values than tagging variants are of direct clinical relevance. Therefore, following GWAS studies, many studies focused on attempting to identify the true causal variant in the loci or nearby regions, which were reported to be associated with phenotypes. Ye et al. conducted a PheWAS for 105 presumed functional (stop-gain and stop-loss) variants. They implemented their PheWAS in 4,235 individuals and validated significant associations in another population of 10,640 individuals. ICD billing codes were used directly as phenotype groups. Two instances ICD codes were required for definition of cases. Codes with less than 8 cases were excluded from analyses. A total of 4,841

coding groups were analysed. Logistic regression models were fitted adjusting for sex, and year of EMR. The association with the lowest *P* value was between rs3731608 and current long-term drug use ($P = 3.2*10^{-6}$). However, it was not replicated in the validation set. In addition, the nonsense variant rs2736911 was found to be nominally associated with age-related macular degeneration ($P = 0.030$). In the validation set, the observed association was directionally consistent, although not statistically significant ($P = 0.081$) (215).

In 2015, Hebbring et al. conducted a text-based PheWAS in 4,235 individuals for 5 SNPs reported by GWAS, including rs3135388, rs9501572, rs12678919, rs220073 and rs1061170. In their study, phenotype groups were defined by clinical texts rather than ICD codes. In brief, a total of 1,564,831 participant clinical notes were extracted, representing 423,537,905 unique words. For each participant, all clinical notes were concatenated together, and broken down into strings of words of four forms, single word, two adjacent words, three adjacent words and four adjacent words. As a result, there were a total of 270,885 single word, 7,507,412 double, 40,568,628 triple and 92,755,314 four word strings. Then these word strings were cross referenced with terms from UMLS. A total of 23,384 word strings appeared in both clinical texts and UMLS, and the dataset was restricted to conditions with at least 50 cases. Then, for every SNP, its associations with all these 23,384 word strings were analysed. Associations were tested by chi square test. All five SNPs were nominally associated with the previously reported phenotypes at $P<0.02$ in the PheWAS. Associations between rs1061170 and word strings 'macular degeneration' ($P = 1.8*10^{-8}$), 'non-exudative' (subtype of macular degeneration, $P = 2.3*10^{-7}$), 'exudative' (subtype of macular degeneration, $P = 1.4*10^{-6}$), and 'visudyne' (commonly used drug in macular degeneration treatment, $P = 3.9*10^{-7}$) survived conservative Bonferroni correction (216).

From 2016, the number of PheWAS began to take off. During year 2016 and 2017, there were 16 PheWAS studies published, which used data from EMR. Among these 16 studies, there were 6 using the 'phecode' system proposed by Denny et al, and 8 studies choosing the ICD holistic way of defining phenotype. In addition, one study

mapped ICD-9 codes to SNOMED-CT codes as Bolland and colleagues did. The last study used the ICD holistic way, but they introduced a novel Bayesian analysis framework and compared results between their Bayesian analysis and general linear models. Details for the methods they employed and their main findings were listed in Table 8 and Table 9.

All above PheWAS were conducted in populations with EMR data. In addition, there are other PheWAS studies whose phenotyping were not based on EMR. Pendergrass et al. conducted a PheWAS for 83 GWAS-derived variants in five population-based samples collected from different sites of multiple ethnicities (70,061 total participants). Their PheWAS were based on epidemiological studies, five studies collected data which included a wide range of common diseases, risk factors, intermediate biomarkers and quantitative traits. Since the phenotype collection was quite independent across individual studies, a total of 105 phenotype-classes were manually created to bin phenotypes from different studies together, and to compare results between individual studies. From their PheWAS, 111 associations were significant from the same ethnicity, SNP and phenotype-class across two or more study sites. Among them, 52 were reported by previous studies, 26 associations represented phenotypes closely related to previously reported associations, and 33 associations were potentially novel. The most significant novel associations were between rs1333049 (*CDKN2A/B*) and haemoglobin levels in African Americans, which had been previously reported to be associated with type 2 diabetes in Europeans (233).

Similarly, Hall et al. conducted another PheWAS on epidemiological samples. Their study covered three surveys of three ethnicities (non-Hispanic whites, n=6,634; non-Hispanic blacks, n=3,458; Mexican Americans, n=3950). A total of 1,008 phenotypes were collected and they were binned into 184 phenotype classes. Associations between 80 GWAS-derived SNPs and phenotypes were tested, stratified by ethnicity. From their PheWAS, 69 associations were significant, of which 39 replicated previously reported association, 9 associations were closely related to previously reported phenotypes and 21 represented novel associations. A total of 13 SNPs showed evidence of pleiotropy (234).

Integrating the idea of PheWAS and MR, Millard and colleagues implemented a MR-PheWAS study of BMI. They created a genetic score comprising 32 BMI-related SNPs, and analysed associations with the score and 172 phenotypic outcomes. As explained before, this equals a two-stage MR. After Bonferroni correction, only the putative association between BMI and HDL-cholesterol at age 9 kept significant. In addition, their study suggested an possible causal association between BMI and a global self-worth score (235).

In order to identify pharmacogenomic associations, Moore et al. conducted a PheWAS for association between 5,954,294 SNPs and 27 laboratory assays in 2,547 pre-treatment individuals from 4 AIDS clinical trial studies. Age, sex, the first 5 ancestral principal components and the square root of CD-4 cell counts were treated as covariates and adjusted in their analyses. From their results, 20 SNP-phenotype pairs matched identical or very close to GWAS catalog reported associations. In addition, 23 SNPs in 29 associations differed considerably from those in the GWAS catalog, including an association between rs10494326 and neutrophil count, and an association between rs2201841 and plasma chloride concentrations (236).

Roesch et al. studied the association between fibroblast growth factor (FGF) 19 and FGF21 and 205 clinical outcomes, including anthropometric, diagnostic, biomarker and medication variables, in 62 patients with type 2 diabetes and 66 controls without type 2 diabetes. From their analyses, 21 variables were associated with either FGF21 or FGF19, but never both. Only the association between high glucose and high FGF21 survived Bonferroni correction ($P = 3.0*10^{-5}$) (237). As a drawback of their study, the sample size is quite limited, which was only 128 in total. It was possible that the associations they observed was caused by chance, but they did not make further effort exploring that.

Furthermore, they applied Wilcoxon Rank sum tests and least-square linear regression in the PheWAS and reported only the *P* value. However, P values do not provide enough information, it could be better if they reported the effect estimates (e.g., coefficient value) of the associations.

In year 2016 and 2017, there were 6 PheWAS studies published, which did not base their phenotype analysis on EMRs. Three of them used trial data as their study population, and the other three study used epidemiological data. Details of their methods and main findings were listed in Table 8 and Table 9.

2.2.2 Critical steps in conducting a PheWAS

Several steps are essential for a PheWAS, and I will describe them in detail in this section. A flow chart on the critical steps of conducting a PheWAS and on how to report results of PheWAS are presented in **Figure 9** and **Figure 10**.

Step 1: Determining the study sample

Any type of study population with comprehensive collection of phenotypes can be used in PheWAS. To explore associations between variants and phenotypes comprehensively, a large population with blood samples and outcome measurements is crucial. As a rich resource for disease prevalence and medical history, an Electronic Medical Record (EMR) based population is the choice for the majority of previous PheWAS studies (199, 200, 202). Most of these PheWAS studies defined their case and control groups by ICD codes within the hospitalization episode data within the EMR system. The advantage of using an EMR cohort is that ICD coding makes phenotyping straight forward. With the hierarchy structure of ICD codes, samples can be categorized into case/control groups in a high throughput and effective manner. In addition to billing code, other information, such as clinical texts in the EMR data can also be integrated to the phenotyping procedure to make phenotyping more accurate and sensitive (216). However, in a general population with EMR data, it is hard to

study some outcomes of low prevalence due to limited power caused by relatively small case numbers.

In addition to the comprehensive PheWAS described above, there is also the targeted PheWAS study design which focuses on only a certain category of outcomes, or biomarkers and other disease related risk factors (247). In that case, the sample does not need to be an EMR population, but samples with the variables of interest are suitable.

Furthermore, some studies may focus on individuals with a specific disease or side effects of a new drug, and then a clinical trial can be a good choice (236, 243). In a trial design, the number of participants is guaranteed and thus the study power can be estimated. In addition, since confounders are well controlled, pharmacogenomics associations can be better explored (236).

In recent years, large biobanks are emerging, such as the UK Biobank, China Kadoorie Biobank (248), the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) (249), and the Veteran Administration's Million Veteran Program (MVP) (250). They provide a very wide spectrum of variables, such as EMR data, serum biomarkers, urine biomarkers, as well as lifestyle and environment exposures. With nearly all the potentially interesting variables, these biobanks could be used to explore associations in multiple levels and resolutions, which could eventually form networks or pathways to uncover biological mechanisms behind statistical associations.

However, as a common challenge, especially for those using prospective cohort linked EMR cohorts, it is hard to study some outcomes of low prevalence due to limited power caused by relatively small case numbers. In addition, as with all epidemiological studies, biases in the sampling process should also be carefully considered and discussed. As with UK Biobank, since it is of higher than average socio-economic status compared to the whole UK population, genotype-phenotype associations which differ between social economic strata might not be well identified.

Step 2: Selecting predictors

The selection of predictors depends on the aim of study. Genetic variants could be involved as predictors. Candidate genetic variants could be determined in different ways. Firstly, variants which were identified by previous studies (e.g., GWAS) to be related with certain outcomes could be used (199, 202, 203, 210). By using GWAS derived variants, the already established associations could be validated and meanwhile, those selected variants could probably be found to be associated with other outcomes not yet found to be relevant by previous studies. Secondly, and more comprehensively, an independent GWAS could be conducted ahead of the subsequent PheWAS, and then significant SNPs in GWAS used as predictors in the following PheWAS (206, 211). This kind of study could help identify new GWAS and PheWAS related variants. Lastly, since most variants of GWAS are just tagging variants without any biological function, fine mapping or functional stop-gain or loss variants could be used (215). For instance, for a certain outcome, the well-established associated locus can be fine mapped first in an attempt to identify causal variants and then these variants can be used in subsequent PheWAS analysis. In addition to nuclear DNA, mtDNA can be considered too (209).

In addition to genetic predictors, non-genetic predictors are also an option. PheWAS can be used to the study the broad associations between biomarkers (or any intermediate phenotypes) and disease outcomes, and possibly identify common biological pathways. For example, autoantibodies (204), enzyme activity (205) have been used by previous PheWAS as predictors.

Step 3: Phenotype coding

In a PheWAS study hundreds of phenotypes are simultaneously analyzed and thus phenotyping is the most important and demanding step in a PheWAS. Several types of phenotyping are possible.

In a study of population linked with EMR, ICD codes are a valuable resource for phenotyping. There are two ways to transform ICD codes into PheWAS phenotype groups, i.e. ICD curated and ICD holistic. A well-known way of curating ICD codes is

the PheWAS code (phecode) system developed by Denny and colleagues (207). In the phecode system, codes that represent common aetiology are combined together in the same phenotype group, while clinically distinct phenotypes which are represented by a single code are divided into separate phenotype groups, such as Type 1 Diabetes and Type 2 Diabetes (199). Controls for a specific phenotype group are all participants that do not have prevalent ICD codes which define cases or highly related codes. An alternative way of curating ICD codes is the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) codes, which is a standardized and comprehensive clinical terms. As what has been done by previous studies, ICD codes were first mapped to SNOMED-CT codes, as they are more clinical relevant, and then associations between predictors and SNOMED-CT code groups were analyzed (212, 217). In a holistic manner, phenotype groups are defined at multiple levels of resolution (for example in an ICD-9 holistic phenotyping, individuals with code 695.11 are defined as a case group, but they belong to the super case groups of 695.1 and 695 as well). Those without recodes of each billing codes are classified as controls in each phenotype group (203).

In addition to the phenotyping rationale, there are some other factors to be considered for data analysis. To increase positive predictive values, individuals are always required to have presentation of a specific code more than twice within the EMRs to be considered as cases. This is especially true for populations from the USA because some codes are recorded as a hypothetical reason for a test, and thus they are not suitable as cases. To increase statistical power and help control the multiple testing burden, case groups with sizes smaller than a certain number (e.g., 20, 25 or 40) are not analyzed. Alternatively, some studies determined a threshold for prevalence of ICD code which will go into data analysis, for example more than 1% (204).

ICD code-based phenotyping is straightforward and thus rapid and effective. However, ICD codes omit some useful information, such as biomarker level or clinical text, which may negatively impact the accuracy of phenotyping. In addition, both curated and holistic phenotyping have their specific advantages and disadvantages. Since multiple ICD codes are clustered into the same phenotype groups, compared with the

holistic approach, the curated phenotyping has smaller number of phenotype groups and thus greater case numbers in each group, which will increase the statistic power and decrease the sample size needed. In holistic phenotyping, the size of some case groups will be too small to be analyzed and the multiple testing burden and the number false positive findings are greater. But holistic coding has the advantage of making no assumptions regarding the genetic or environmental contributions to any one disease (251), which is similar to the design of GWAS studies, which is a hypothesis generating design. In addition, in a statistically significant association found by curated coding, the signal might be led by one of the ICD codes which are binned into the same phenotype group. In such a case, analyzing individual ICD codes separately in the holistic manner could help in clarifying potential biologically causal pathways.

The majority of previous PheWAS were conducted with the ICD 9th version. The implementation of ICD 10th version, has resulted in more and more hospitalization data having to be adapted to this new system, including data from the UK Biobank. Whereas the ICD 9 system allows for nearly 17 000 possible codes, the ICD 10 system allows for more than 155 000 different codes (251). In such an updated ICD system with fine granularity, the granularity of phenotyping algorithm should be considered carefully based on sample size and disease prevalence.

Apart from ICD code based phenotyping, there are some other efficient options. Since clinical texts present more information than ICD codes, they could be considered in phenotyping. Hebbring et al. conducted such a study exploiting text-based phenotyping (216). In their study, clinical notes of participants were broken down into unigrams (one word), bigrams (two adjacent words), trigrams (three adjacent words) and 4-grams (four adjacent words), which were subsequently cross referenced with the National Library of Medicine's Unified Medical Language System (UMLS) medical dictionary. Word strings which were observed in both the UMLS medical dictionary and clinical text data and meanwhile had at least 50 cases were used to define phenotype groups. By comparison with ICD based coding, Hebbring demonstrated efficacy of text-based phenotyping. But as their phenotyping on word strings was imperfect, they were still faced with the problem of misclassification. Introducing

natural language processing would be helpful in refining this method. In the future, an advanced phenotyping methodology, which combines ICD codes, laboratory values, medication records, clinical texts, imaging data or other data, would be expected. Denny et al. conducted a combination study of GWAS and PheWAS on hypothyroidism, and in the preliminary GWAS study, an algorithm incorporating ICD9 codes, laboratory values, text queries, and medication records was exploited for identifying cases/controls of hypothyroidism. In the subsequent PheWAS, phenotyping was still relied on ICD codes (200). Up to date, there is still no PheWAS approach which incorporates all above information in phenome-wide phenotyping.

There are some epidemiological survey samples which do not have participants' hospitalization data available (233, 234). However, with the great number of phenotypes collected, a PheWAS could still be feasible on this kind of data. The problem is that phenotype definition in epidemiological studies is not standardized. Different surveys will have their own phenotype variables and number of phenotypes. Therefore, for collaboration between study sites to gain a greater sample size and power, or to validate and compare results between study sites, phenotypes must be manually binned into phenotype classes. Then, a certain threshold (e.g. $P<0.01$) observed in two or more studies for the same predictor, phenotype class and race/ethnicity and consistent direction of effect could be considered as significant. As a tool for phenotype harmonization, the DataSchema and Harmonization Platfrom for Epidemiological Research (DataSHaPER) has been developed and its utility has been demonstrated (252).

As discussed above, the prevalence of disease is not the only outcome that PheWAS are interested in. Some other phenotypes, such as body mass index, total cholesterol, high-density lipoprotein cholesterol or other laboratory phenotypes could also be involved in a PheWAS study (236).

Step 4: Statistical analysis

To establish the associations between predictors and multiple outcomes, logistic or linear regression can be used under an additive model, in which genetic predisposition

of wild type, heterozygotes and effect allele homozygotes are defined as 0, 1 and 2 respectively. Meanwhile, chi-square and student T-test can also be considered for use. For small size groups (n=5), Fisher's exact test needs to be conducted. Possible confounding factors include age, sex, ethnicity, year of EMR records, study site, and ancestral principle components (calculated on genotype information to adjust for population stratification), which is commonly the first three or first five.

Common software used in this type of analysis include R, PLINK and SAS that can all complete the high throughput analysis effectively. In addition, EIGENSTRAT and EIGENSOFT are helpful in generating ancestral principle components and QUANTO can be used for sample size and power calculation.

Similar to genome wide association studies, PheWAS is also challenged by multiple comparison testing. Four main ways could be used to correct the observed *P*-values for multiple testing. These include Bonferroni correction, false discovery rate (FDR), permutation, and inter-database replication. Since phenotype groups in PheWAS are correlated, especially in a holistic ICD coding based PheWAS, Bonferroni is widely considered to be over-conservative. The FDR is the expected proportion of erroneous rejections among all rejections. The more hypotheses are rejected, the more errors are treated as acceptable (253). Thus, FDR is more tolerable than Bonferroni and has been proven effective in dependency situations. Permutation, which could be implemented with PLINK, takes the correlations between phenotypes into consideration and thus could be treated as another option. Inter-database replication needs two or more comparable and independent cohorts. If an association is statistically significant in more than one database with consistent effect direction, it is considered as positive (233, 234, 236).

Recently, a new Bayesian analysis framework has been developed by Cortes and colleagues, in order to deal with the interrelations between phenotypes. In this analysis, the relationships between phenotypes were accounted by the Markov process. It calculated a posterior probability, and do not need to implement the multiple testing correction. This method was proved to increase power, meanwhile identify novel

associations which could not be found by conventional analysis methods (226).

Step 5: Interpreting significant results

In a PheWAS, if the SNPs or other predictors were determined by GWAS or other previous studies, results replicating known associations are expected. For the replicated association, interpretation combining previous studies and known functional and biological information is suggested. It is relatively straightforward to identify true associations when knowledge of expected associations is incorporated into the interpretation of PheWAS results (215). In contrast, some known associations may not be replicated as expected. This inconsistency could result for several reasons, such as small case number due to low disease prevalence and small sample size, different environmental exposures and genetic/phenotypic heterogeneity. In addition, in an ICD coding based PheWAS, the phenotyping algorithm might also drive this. In the PheWAS conducted by Denny and colleagues in 2010, a chart view for systemic lupus erythematosus (SLE) cases defined by ICD9 based algorithm was presented. Among the 141 cases defined automatically by the algorithm, only 95 (67%) had documented or probable SLE as indicated by their treating physicians. The other false positive records contained SLE ICD9 codes due to tests or hypothetical diagnoses which were later dismissed (199). In addition, exclusionary events predating electronic records or events occurring outside medical facilities could also cause further bias in classification (200). Therefore, for the results which are not replicated but of major importance and with strong functional evidence, a manual inspection of the automatically defined cases is recommended.

In a PheWAS, a single predictor can be found to be associated with multiple outcomes. In such a scenario of probable evidence of pleiotropy (i.e. a single predictor is associated with more than one distinct phenotypes), the correlations between these outcomes should be carefully considered and discussed. The relationship might be due to a common biological process with known genetic contribution (true pleiotropy). Alternatively, there may be a network or intermediary relationship (false pleiotropy), in which genetic variation may impact the variation of a single phenotype, but variation in that phenotype could then results in changes in other downstream phenotypes

indirectly (234). Namjou found *PTPN22* to be associated with JRA, T1DM and thyroiditis; pleotropic effects which were expected due to known underlying biological correlations (210). *FTO* variants were found to be associated with T2D and obesity by Cronin et al., and they further discussed the relationship between *FTO*, T2D and BMI (208). A mediation analysis will be helpful in clarifying this kind of relationship.

In addition, combining separate PheWAS results into a network could be a way to understand the underlying etiology and find true causation. For example, Hall and colleagues found rs328 (associated with HDL Cholesterol) near *LPL* and rs174547 (associated with ferritin) near *FADS1* both to be involved in TGF-beta receptor regulated NetPath pathway (234).

Step 6: Reporting results

On reporting PheWAS results, there are several helpful suggestions to be followed (Figure 10). The study sample, sample size, ethnical structure, other demographical data, outcomes collected, and whether it is a general population sample or clinical sample should be reported first.

Then the way of selecting targeted predictors should be explained. If they were selected based on known associations, the previous findings should be included to be compared and discussed further in later sections.

A crucial part is the rationale of phenotyping and being explicit about how this was conducted. The number of phenotype groups and definition of cases and controls should be reported. The sensitivity, specificity and positive predictive value of the phenotyping method, if applicable, could be reported. If outcome aggregation is involved, the method of binning should be explained (examples of some of the phenotype groups would be helpful). The problem of misclassification should be explored and reported. In addition, for curated ICD code phenotyping, the application of the rule of two (an individual is required to have at least two presentations of the same code to be considered as a case) and the minimal case number of phenotype groups which went to analysis (groups with few cases may be omitted) should also be

explained.

In statistical analysis, the analytic method used should be reported. Details of the covariates included in the analysis should be reported. The way to deal with multiple testing challenge or the corrected significant threshold should be reported.

The results of a PheWAS could be categorized into replicated results, related results (results of predictors associated with an outcome which is closely related to another previous reported outcome) and novel results. If there are some interesting predictors, which show some evidence of pleiotropy, they can be placed into a separate table or figures.

For a PheWAS, the result table should include predictor information (e.g. SNP ID, SNP location, nearest gene), previous reported phenotype (to make comparison between previous studies), *P* value, effect size, and case numbers of each phenotype, which varies between different phenotype groups in PheWAS.

A Manhattan plot can also be useful for the result presentation. In contrast to that of GWAS, the X axis of Manhattan plot of PheWAS is the phenome spectrum. Lines of different significant levels could be plotted. And besides, the effect size could be presented with the size of dots. An alternative option is sun plot (234, 254). In a sun plot, the results across the whole phenome spectrum cannot be shown. However, the statistically significant results for a specific predictor could be more easily visualized. With the name of the predictor labelled within a circle in the center, every significant result is presented surrounding it, whose significant level is presented by the length of a line connected the circle and the phenotype names around.

To explore relationships or common biological pathways between different predictors and phenotypes, results are also suggested to be presented into a network or pathway manner.

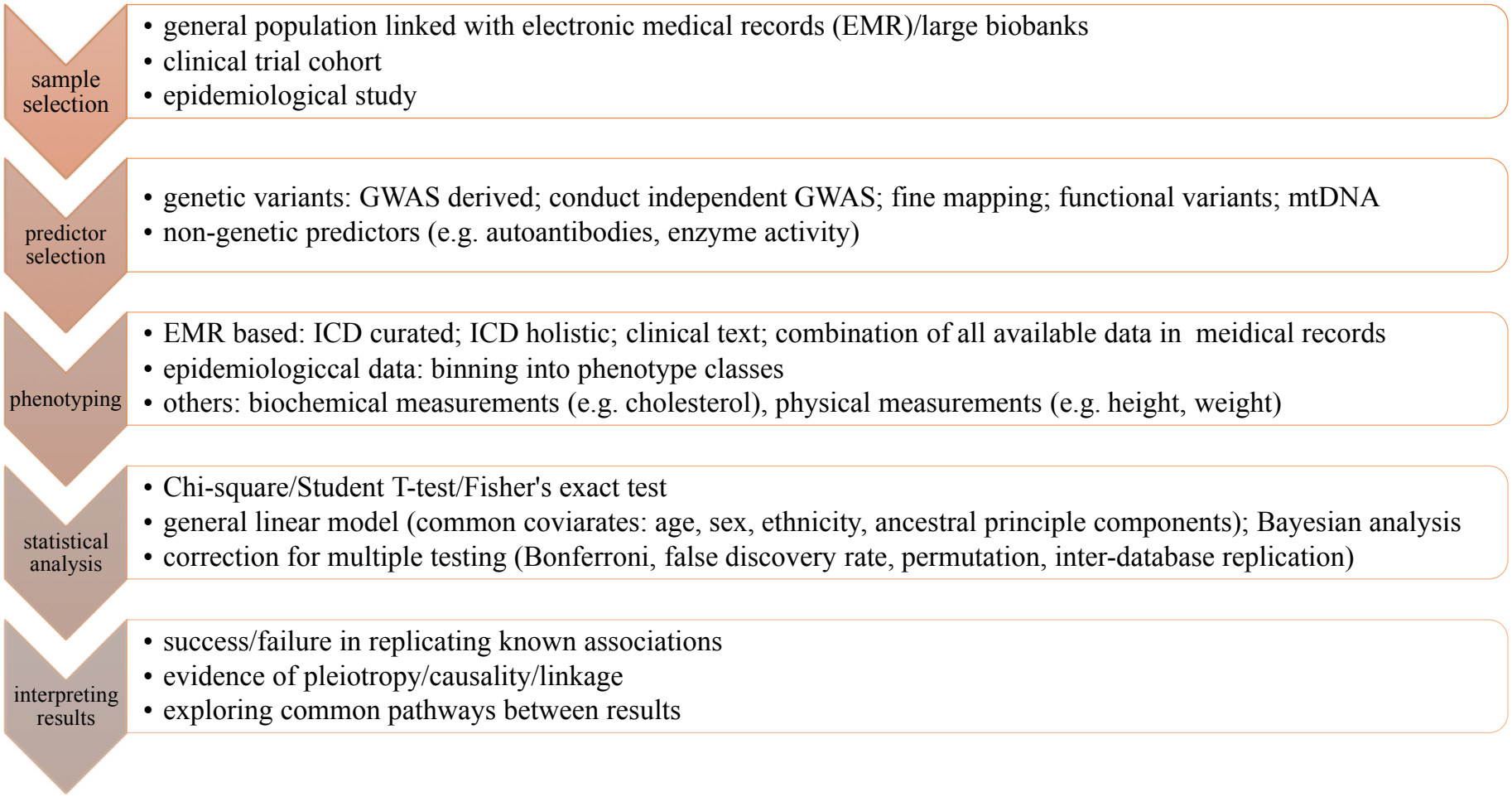


Figure 9. Flow chart of conducting a Phenome wide association study.

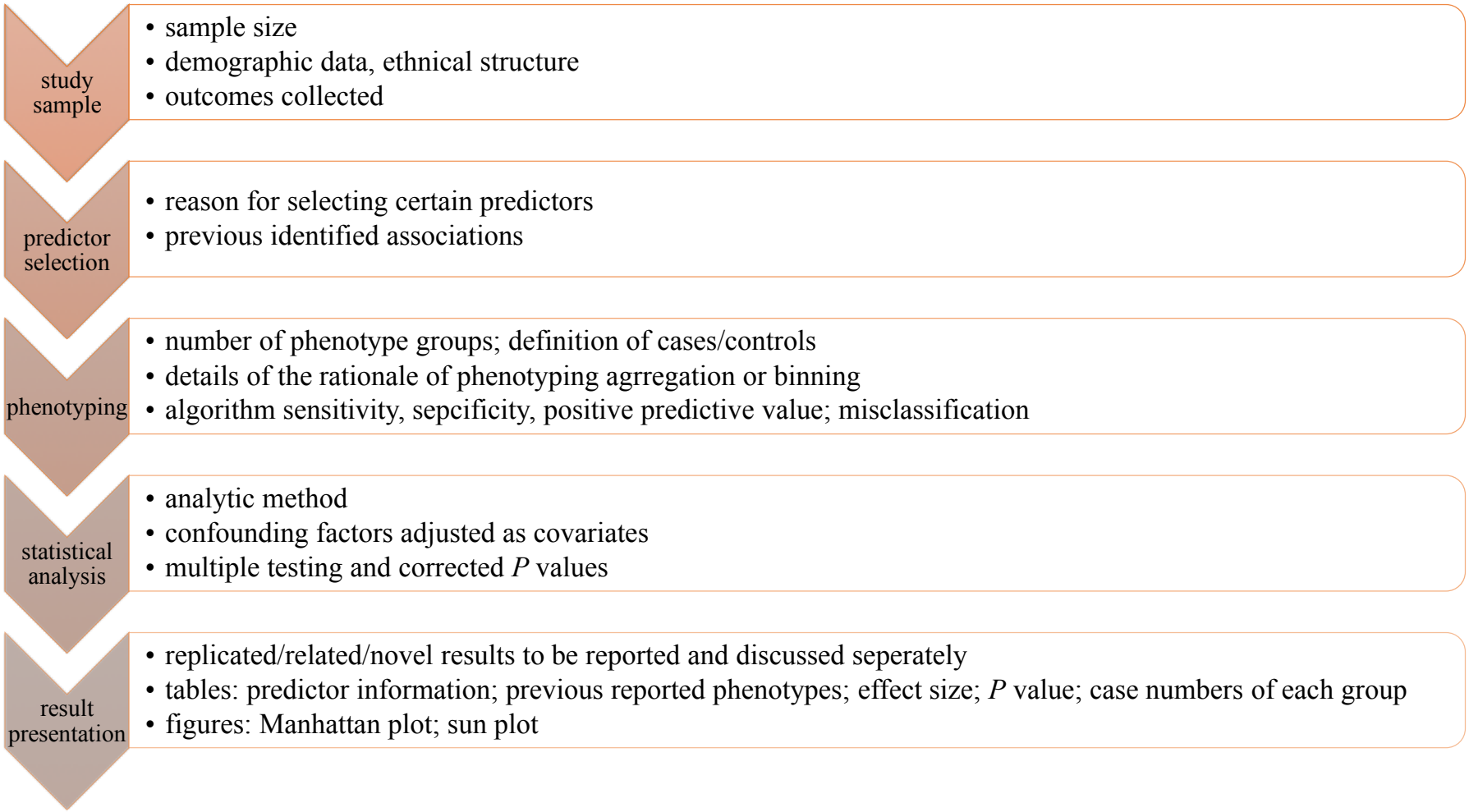


Figure 10. Flow chart of reporting phenome wide association study results.

2.2.3 Follow up studies for significant associations from PheWAS

Since PheWAS is a hypothesis-generating methodology similarly to a GWAS, statistically significant results from a PheWAS need to be replicated. Generally, an independent case-control study would be an option for replication, which could avoid the multiple testing and misclassification biases of the high throughput PheWAS. To determine the sample size of disease specific replications, assumption on prevalence, allele frequency and effect size must be considered. In addition, in the separate replicating study, defining cases and controls must be careful and precise, with preference to a combination of all available information such as serum biomarker, trait measurements, gold standard diagnosis, and even a case by case manual adjudication. In addition, possible effects of covariates such as age, sex, and environmental exposures on the association between predictor and phenotypic outcome should also be investigated (233).

After a result has been replicated by independent studies of different type (i.e., general population, case-control, prospective cohort or trial), further follow up analysis could be conducted. Since most predictors of PheWAS are SNPs derived from published GWAS, PheWAS results have the same problem as those of GWAS. Many SNPs are just tagging SNPs in intergenic region without any function. Thus, fine mapping searching for causal variants in nearby or located genes is a critical step. Subsequently, functional or pathway studies could be conducted to reveal the true biological mechanisms underlying the significant results, which will pave the way for future genetic prediction or drug development.

2.2.4 Strengths and limitations of PheWAS

Strengths

Focusing on SNPs identified by previous GWAS, PheWAS provides an approach for replicating GWAS results. In addition to replication and validation, PheWAS could also help in discovering novel genotype-phenotype relationships which are not found by disease specific GWAS or candidate gene studies. In many cases, the initial association found by GWAS may not be due to causality. By exploring associations between specific SNPs and a wide spectrum of outcomes, therefore a PheWAS could

help in identifying potentially causative associations underlying discoveries in disease-specific GWAS. If conducted in combination with GWAS, a network between genotypes and phenotypes can be constructed, providing a comprehensive catalogue of human diseases associated with published variants and a broad insight into the pathophysiology of multiple disease processes.

The majority of PheWAS studies are conducted in populations with EMR data. This brings several certain advantages. Firstly, since individuals commonly experience multiple health conditions which will be all recorded in EMR, there is potential to support investigation of a wide variety of diseases. Meanwhile, the logical architecture of ICD codes makes phenotyping relatively straightforward and cost effective. In addition, since records of individuals are recorded longitudinally in EMR, it holds the promise of examining longitudinal healthcare outcomes such as disease complications or response to drug therapies, which has been demonstrated in a previous PheWAS by Ritchie and colleagues (206).

Moreover, in pharmacogenomics, by exploring associations between genotype and drug targets as well as other outcomes or traits, PheWAS could provide an estimate of drug efficacy and toxicity, which will also help in identifying drug side effects.

Limitations and future developments

As a high-throughput analysis approach, PheWAS is faced with the problem of sample size. In previous PheWAS studies, some prior associations were successfully validated, however, others were not. Failure in replicating an identified association may be a result of limited power. The power to detect an association is determined by the minor allele frequency, effect size and prevalence. For some case groups, due to limited sample size and low prevalence, case numbers may be too low to draw any significant conclusions. This can also happen for rare variants. In addition, as a result of high-throughput analysis, only standard covariates (i.e., age, gender, and ethnicity) are usually adjusted for in the analysis. However, other important confounding factors, such as environmental exposures and family history, are not considered. To resolve the problems due to limited sample size, large samples with genetic data linked to EMR

records are needed, which will increase the power of both GWAS and PheWAS studies and pave the way for studies on rare variants and rare diseases. In the near future, some large population cohorts will come into being (e.g., Kaiser Permanente, the Million Veterans Programs, the UK Biobank) and that will benefit future studies.

Since SNPs identified by past GWAS are mostly used in PheWAS, we still have the same problem of identifying causal variants. Most SNPs identified by GWAS are just intergenic ones which do not have any biological function and they are just markers for other casual variants in linkage disequilibrium. As proved by Ye et al. (215), future studies targeting functional variants (e.g., functional stop-gain and stop-loss variants) or deleterious variants will be more effective in finding causal variants and biological mechanisms and at the same time are likely to have a relatively higher effect size, which may also increase power and mitigate the challenge brought by low allele frequencies.

A major challenge with PheWAS is the multiple testing burden. Common methods for correcting the significance level include Bonferroni, FDR, inter-data replication and permutation. The hypothesis of Bonferroni requires independence between the regressions. However, in a PheWAS study, there is large correlation between phenotypes, which makes Bonferroni overly strict and inappropriate in this situation. Future novel correction methods (e.g., the Bayesian analysis framework) should be expected, which can take the complex relation network between genotypes and phenotypes into consideration.

Developed by Denny et al., an automatic phenotyping algorithm based on ICD9 codes is widely used in PheWAS. However, there are several problems with this ICD coding system. As has been elucidated by Denny et al. (199), there can be false positives in ICD coding, in which billing codes are recorded as a hypothetical reason for a test. Besides, the ICD codes can miss medical conditions that predate an EMR system. Therefore, the sensitivity and accuracy of ICD coding is imperfect and it may not provide a comprehensive assessment for individuals' disease history. The replicability of phenotyping in previous PheWAS maybe limited partly due to this reason. However,

since multiple phenotyping methods are available (phecode, SNOMED-CT, ICD holistic, epidemiological measurements) and there are several options to choose in process of phenotyping (digits of ICD codes to consider, use of the rule of 2, least number of cases to involve), different case-control groups could be generated by different choice of phenotyping methods. Methodological choices are determined by study aim and available resources. Differences in results generated by different phenotyping methods may not necessarily mean low replicability. It could be due to differences in statistical power of different methods, or different phenotype groups defined by different methods. Hence, the replicability of phenotyping process is not straightforward to investigate. In the future, some advanced phenotyping algorithm incorporating of various available information such as laboratory data, medication text, imaging data as well as ICD coding and taking advantage of natural language processing and machine learning should be established to provide a reliable automatic phenotyping tool of high sensitivity and high positive predictive value. Besides, this improved phenotyping algorithm is also expected to improve the replicability of PheWAS phenotyping.

Finally, samples within an EMR system are always clinical populations which are exposed to a healthcare system and could have different characteristics from the general population, such as age or prevalence of diseases. Thus, associations found from EMR populations might not apply to the general population well. To resolve this concern, medical care systems covering a much broader population and richer resources of information which should also include daily and home care data as well as data in clinical facilities, and combination of EMR populations across a variety of geographical locations can be a possible solution. This is especially true for my study, since only the inpatient ICD data has been released by the UK Biobank. For the health care system in UK, a large proportion of medical care episodes were in general practice and outpatient departments. With data linked with inpatient EMR only, cases treated only by general practice and outpatient departments will be misclassified as controls. The degree of misclassification differs by outcomes. For outcomes which normally go to inpatient department, this problem is minimal, such as cancer and cardiovascular outcomes. However, for outcomes which are mostly treated out of inpatient

departments, such as hypertension, depression and vitamin D deficiency, the problem of misclassification will be more prominent.

Chapter III: Aims and Objectives

In Chapter I, I described that vitamin D deficiency is a highly prevalent condition and that it has been found to be broadly associated with many health outcomes. However, there is a lack of evidence to support a causal role of vitamin D. In this chapter, the aims and objectives of the thesis will be presented. In short, this thesis has two major aims, to investigate the association between vitamin D related SNPs and a wide range of health-related outcomes, and to explore the causal relationship between vitamin D and health-related outcomes.

3.1 Aim 1: Association between vitamin D related SNPs and health-related outcomes in the UK Biobank cohort

The first aim of my thesis is to present the results of a number of PheWAS analyses conducted with data from the UK Biobank cohort. Since 6 SNPs have been found to be associated with vitamin D in previous large GWAS studies in white populations, I aim to explore the phenome wide associations of these SNPs. Specific objectives include:

1) To conduct a systematic literature review on existing PheWAS studies, their applied methodology and most important findings; discuss their advantages and challenges, and try to raise some suggestions for future PheWAS methodology. (presented in Chapter II)
2) For the 6 vitamin D related SNPs reported by the largest GWAS, describe their genotype distribution in UK Biobank cohort, and test for their HWE. For each SNP, test the associations with confounding factors, including UK Biobank assessment centre, gender, age, outdoor activities in winter/summer, qualifications, household income and drinking and smoking status.
3) Prepare and code the relevant phenotype data (especially the inpatient data, cancer register and death register data) for use in the PheWAS analysis.
4) For each SNP and scores of all SNPs (score weighted by their effects estimated by previous GWAS), conduct PheWAS to comprehensively test their associations with all medical outcomes defined by ICD coding system.

3.2 Aim 2: Investigate the evidence for causality between vitamin D and health outcomes in the UK Biobank cohort using a MR methodology

Following the first aim, a MR methodology will be used to explore the causal association between vitamin D and health outcomes with vitamin D related SNPs in aim 1 as IVs. Specific objectives include:

1) Conduct a systematic literature review for all previous MR studies on vitamin D, to explore the findings of previous MR vitamin D studies (presented in Chapter I, section 1.4).
2) In UK Biobank, conduct MR analyses for PheWAS outcomes that survived multiple testing correction, for outcomes with sufficient power, and for outcomes which had been explore by previous MR studies. Multiple MR methods will be used in this stage, including two-stage method, Inverse Variance Weighted (IVW) MR and Egger's MR, followed by sensitivity analyses.

Chapter IV: Methodology

In this chapter, I will describe the methodology of my study, which comprise two parts. In the first part, I will give the details of UK Biobank cohort, which is the population I will use in my study. In the second part, I will give details of the statistical methods which were explicitly used in my study.

4.1 The UK Biobank Cohort

In this section, information on the UK Biobank cohort, which are relevant to my study will be given. This include the study design, ethical approval, participant recruitment, data collection and release of UK Biobank data. Since UK Biobank is publically available, all details of this cohort are online at: http://www.ukbiobank.ac.uk/. For my study, an application for data usage was submitted to the UK Biobank group (Application ID: 10775). All the analyses I have conducted are included within this data application.

4.1.1 Study design

UK Biobank is a very large, prospective cohort study founded by the Wellcome Trust medical charity, the Medical Research Council, the Department of Health, Scottish Government and the Northwest Regional Development Agency. It aims to allow detailed investigations of genetic and non-genetic determinants of diseases of middle and old ages, which are commonly caused by a combination of lifestyle, environmental, and genomic factors, with individually modest effects and complex interactions (255). The cohort is designed to include ~500k general population from the UK. It intends to collect comprehensive baseline measurements and phenotypes from every participant, including whole and saliva sample collection, and links participants to their medical records in the National Healthcare Services (NHS) system. The participant recruitment and baseline data collection had been completed during 2006-2010. The follow-up will last for 20 years to facilitate the investigation for a wide range of diseases and outcomes.

4.1.1.1 Ethical approval

In order to keep independent of the UK Biobank team, the UK Biobank Ethics and

Governance Council (EGC) was established by the Medical Research Council and the Wellcome Trust. EGC is responsible for the setup the ethical standards for UK Biobank, monitoring and reporting publicly on the conformity of the UK Biobank project, and advising on the interests of participants and the general public in relation to UK Biobank.

The UK Biobank project has been approved by the North West Multicentre Research Ethics Committee (MREC), which covers the whole UK. It also has got ethical approval from the National Information Governance Board for Health & Social Care (NIGB) in England and Wales and from the Community Health Index Advisory Group (CHIAG) in Scotland. In addition, a generic Research Tissue Banc (RTB) approval (meaning that , research applications from the great majority of researchers will not need separate ethical approval). At lasts, UK Biobank also possesses a Human Tissue Authority (HTA) license. Therefore, researcher receiving samples from UK Biobank do not need a separate HTA license.

4.1.1.2 Participant recruitment and enrolment

UK Biobank identified potential eligible individuals' name, address, sex, date of birth, NHS/CHI number from the NHS records. Invitations were sent to their addresses on NHS system. From year 2006 to 2010, UK Biobank recruited 502,665 UK residents aged between 40-69 years across the whole UK. Consent was given by all participants. They have been informed that they may be re-contacted by UK Biobank (e.g., for a follow-up visit, to seek consent to proposed new uses, etc.). It was explained to participants that they can withdraw their participation in the following 3 forms at any point of time and without giving a reason: 1) no further contact (UK Biobank would no longer contact the participant, but keep the right of obtaining further information from the health-relevant records, using their data and biological samples); 2) no further access (UK Biobank would no longer contact the participant, would not access their health records, but would still use the information the biological sample that have been collected); 3) no further use (totally quit, any data of this participant would not be used in any UK Biobank research project).

4.1.2 UK Biobank baseline data collected at recruitment

Participants who agreed to join were invited to an initial assessment visit, which happened in one of the 22 assessment centres across the whole UK (**Table 10**). The assessment visit comprised electronic signed consent; a self-completed touch-screen questionnaire; a brief computer-assisted interview; physical and functional measurements; and collection of blood, urine, and saliva samples (255). The questionnaire and interview cover participants' family history and early life exposures; psychosocial factors; environmental factors; lifestyle information; health status; hearing threshold and cognitive functions. The physical measurements include blood pressure and heart rate, grip strength, anthropometric measures, spirometry, bone density, arterial stiffness, eye examination and fitness test. Further detailed descriptions of the collection of phenotypes/information from UK Biobank relevant to this study are given below.

4.1.2.1 Phenotypes collected from touch-screen questionnaire

A touchscreen questionnaire was completed by every participant at recruitment, with a variety of demographic, lifestyle and other phenotypes collected, including gender, date of birth, height, weight, time spend outdoors in summer/winter, average household income before tax, qualifications, alcohol intake frequency, participant's home address and self-reported medical conditions (for details of the touch-screen questionnaire: http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=113241).

4.1.2.2 Age and gender collection and presentation in UK Biobank data

Participants' date of birth, month of birth, year of birth are recorded by UK Biobank into 3 distinct variables. Date of birth is highly sensitive and is supplied by UK Biobank to researchers only when necessary. Month of birth and year of birth are readily available to researchers. In addition, age at recruitment and age when attending assessment centre are also derived by UK Biobank and ready to use for researchers. In my study, I used age when attending assessment centre as a covariate in analyses.

Gender was reported by participants at assessment centre and was used as a covariate in my data analyses. There are 229,165 males (45.6%) and 273,455 females (54.4%)

in the cohort.

4.1.2.3 Height, Weight and BMI collection and presentation in UK Biobank data

Standing height was measured through assessment using a Seca 202 device, with a mean of 168.68 cm and SD of 9.27. Weight was measured by a variety of means during initial assessment, and then merged into a single item, with a mean of 77.97 kg and a SD of 15.92. Body mass index (BMI) was derived by weight in kilograms divided by square of height in metres, with a mean of 27.39 kg/m^2 and a SD of 4.79.

4.1.2.4 Other baseline variable collection and presentation in UK Biobank data

Outdoor activity time was reported as the number of hours participants spend outdoors on a typical summer/winter day. The home address of participants was represented by the east co-ordinate and north co-ordinate of the postcodes to which invitations were sent, using the Ordnance Survey reference (256). Household income before tax, qualifications, alcohol intake and smoking status were reported as categorical variables (**Table 10**).

Table 10. Characteristics of baseline phenotypes.

Variable Name	Feature	
Continuous	Mean (S.D.)	Number of missingness
Age at assessment	56.85 (8.13) years	0
Standing height	168.48 (9.27) cm	2,490
Weight	77.97 (15.92) kg	2,716
BMI	27.39(4.79) kg/m^2	3,041
Time spend outdoors in summer	3.93 (2.33) hours/day	3,997
Time spend outdoors in winter	2.22 (1.84) hours/day	3,997
Categorical	Levels	Number of

		participants
Sex	Male	229,165
	Female	273,455
	Missing	0
Household income before tax	Less than £18,000	102,711
	£18,000 to £30,999	118,413
	£31,000 to £51,999	120,677
	£52,000 to £100,000	93,032
	Greater than £100,000	24,645
	Do not know	22,409
	Prefer not to answer	52,085
	Missing	5,353
Qualifications	College or University degree	178,083
	A levels/AS levels or equivalent	143,300
	O levels/GCSEs or equivalent	244,107
	CSEs or equivalent	69,024
	NVQ or HND or HNC or equivalent	98,636
	Other professional qualifications (eg: nursing, teaching)	152,598
	None of the above	88,180
	Prefer not to answer	5,618
	Missing	3,985
Alcohol intake frequency	Daily or almost daily	108,494
	3 or 4 times a week	125,562
	1 or 2 times a week	139,209
	1 to 3 times a month	60,204
	Special occasions only	61,927
	Never	43,135
	Prefer not to answer	612
	Missing	889
Smoking status	Current	54,639
	Previous	186,131

	Never	296,226
	Prefer not to answer	2,153
	Missing	883
Tobacco smoking	Smokes on most or all days	2,873
	Occasionally	1,924
	Ex-smoker	44,012
	Never smoked	71,867
	Prefer not to answer	615
	Missing	381,329
Assessment centre	Barts	12,583
	Birmingham	25,503
	Bristol	43,015
	Bury	28,336
	Cardiff	17,882
	Croydon	27,385
	Edinburgh	17,201
	Glasgow	18,651
	Hounslow	28,879
	Leeds	44,209
	Liverpool	32,818
	Manchester	13,940
	Middlesborough	21,289
	Newcastle	37,008
	Nottingham	33,877
	Oxford	14,062
	Reading	29,417
	Sheffield	30,397
	Stockport(pilot)	3,798
	Stoke	19,440
	Swansea	2,281
	Wrexham	649
	Missing	0

4.1.2.5 Medical and drug history at baseline collection

In addition, with the touch screen questionnaire, participants were asked about their personal medical history for specific diseases (including vascular/heart problems, blood clot/DVT/bronchitis/emphysema/asthma/rhinitis/eczema/allergy, diabetes, gestational diabetes, cancer, fracture/broken bones, other serious medical condition/ disability; url: http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100044) and the age at diagnosis. Following the questionnaire, there was a verbal interview by a trained research nurse on all past and current medical conditions. If the participant had ticked any kind of major disease in the questionnaire, the research nurse confirmed that by interviewing the participant. If in the interview, it turned out that these have been incorrectly selected by participants, the nurse would cancel the wrong response(s). If the participant stated that he/she had other serious medical condition or disability, the nurse would fill in the relevant disease code(s) as coded by UK Biobank data coding 6, which is a customized disease coding system developed by the UK Biobank (different from the ICD coding system, which was used in the health records from linkage to NHS system, for more details on this data coding 6, please see: http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002). If the participant stated in the questionnaire that they had no major illnesses or disability or were not sure, this question was asked again and confirmed by the research nurse.

Similarly, participants were asked about any prescription, medication, or dietary vitamins/supplements they regularly took when completing the touchscreen questionnaire (http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100045). Then during the subsequent verbal interview with the research nurse, any regularly drugs taken were confirmed (http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100075).

4.1.3 Hospital Episode Statistics, Cancer Registry and Death Registry data

UK Biobank participants were linked to electronical medical records (EMR) data, including Hospital Episode Statistics (HES), cancer registry and death registry data. In 2018, UK Biobank participants will be further linked to their General Practice data and imaging data.

4.1.3.1 Hospital Episode Statistics [HES]

HES data consists of three parts: in-patient admissions data, outpatient data and accident and emergency data. Only the inpatient data (including maternity and psychiatric episodes of care) have been released by UK biobank, with the other two parts to be made available in the future. The hospital inpatient data are sent to UK Biobank on an annual basis and includes data on admissions and discharges, diagnostic and operation codes and maternity and psychiatric data. Records from both National Health Service (NHS)-funded and private health care within NHS hospitals were included. The World Health Organization's ICD codes were used to record diagnosis information, and from 1997 onwards, episodes are coded in the ICD 10th version. Any earlier records were coded using the ICD 9th version.

The inpatient data are constructed on an episode basis, and every episode of a participant is a single line of record in the data, which is different to the way UK Biobank baseline data are organised (which is on an individual basis with every row of data representing a unique participant and with phenotypes running across the columns). Episode is defined as a continuous period of admitted patient care administrated under one consultant within one healthcare provider. If the patient is transferred to another consultant or to a different healthcare provider during a spell of treatment, a new episode is generated. A hospital spell is the total time a patient is in hospital, from date of admission to date of discharge. A hospital spell can consist of multiple consultant episodes (e.g., if the patient is transferred between consultants or providers during their time in hospital). Therefore, there can be one or more episodes within a hospital spell; likewise, there can be one or more hospital spells per participant (http://biobank.ctsu.ox.ac.uk/showcase/docs/HospitalEpisodeStatistics.pdf) (257).

The inpatient data are organised into 5 sub-categories, which include admissions/discharge, operations/procedures, diagnoses, maternity and psychiatric admissions. Admissions and discharge include data on admission and discharge, hospital stay and type of care given. The operations/procedures domain contains operation and procedure codes coded by the OPCS. The diagnosis domain contains diagnosis codes for main and secondary diagnoses coded by the ICD system. The maternity part contains data related to maternity inpatient care (e.g., method of delivery, place, sex and status of baby). Finally, the psychiatric part includes data related to psychiatric inpatient care (e.g., carer support, history of psychiatric care, legal status).

4.1.3.2 Cancer and Death Registry data

The cancer registry data are sent to UK Biobank by the Medical Research Information Service, which is based on the NHS Information Centre (for participants residing in England and Wales) and Information Services Division, part of NHS Scotland (for participants residing in Scotland). Cancer registries acquire information on cancer diagnoses from multiple sources including hospitals, cancer centres and treatment centres, hospices and nursing homes, private hospitals, cancer screening programmes, other cancer registries, general practices, death certificates, HES, and Cancer Waiting Time data. The cancer registry data include date of cancer diagnosis, age at cancer diagnosis, type of cancer (coded by ICD), reported occurrences of cancer, histology code and behaviour code (http://biobank.ctsu.ox.ac.uk/crystal/docs/CancerLinkage.pdf) (258). UK Biobank has cancer register data prior to study inception, during the study and after the study finished. The type of cancer is coded according to the version of ICD coding that was relevant for that time period. There is a small amount of records in ICD8, which have been grouped into the ICD-9 tree structure. From year 2000, the type of cancer is coded with the ICD10 structure. The histology and behaviour codes of cancer are represented as separate variables, which are presented as five-digit codes in ICD10-O-3, ranging from M-8000/0 to M-9989/3. The first four digits code the histology and the fifth digit codes the behaviour.

Finally, death registry data are also available in UK Biobank. NHS Information Centre and the NHS Central Register, Scotland updates death certificates for UK Biobank on a quarterly basis. The data includes date of death, age at death, underlying cause of death (only one), contributory causes of death (could be multiple) and description of cause of death (http://biobank.ctsu.ox.ac.uk/crystal/docs/DeathLinkage.pdf) (259).

4.1.4 Genotyping and imputation data

4.1.4.1 Genotyping and imputation

Two 10mL blood sample per participant were collected by UK Biobank per participant. DNA was extracted from 850 μL buffy coat recovered from one of the 10 mL EDTA whole blood sample by a customized and automatic DNA extraction system. Genotyping plates (96 wells with two controls on each plate), 90 μL DNA in each well were shipped from UK Biobank, Stockport, UK to Affymetrix Research Services Laboratory, Santa Clara, CA, USA for genotyping.

DNA samples were genotyped using either the Affymetrix UK BiLEVE Axiom (50,000 individuals) or Affymetrix UK Biobank Axiom (450,000 participants). These platforms are similar with 95% shared contents. UK BiLEVE is a study on chronic obstructive pulmonary disease, which included 50,008 UK Biobank participants with middle or extreme values (relative to the population distribution) of forced expiratory volume in 1 second from the UK Biobank 500k participants (260). The UK BiLEVE Axiom array (807,411 markers) was designed to a) measure rare coding variants, b) optimize imputation performance for common (MAF>5%) and low frequency (MAF 1-5%) non-genotyped variants, and c) cover genes and genomic regions with established or putative roles in lung related outcomes comprehensively. The UK BiLEVE study was completed before the genotyping of 450k UK Biobank participants, and the UK BiLEVE Axiom array was used to finalise the design of the Affymetrix UK Biobank Axiom array. The UK Biobank Axiom array includes 820,967 genetic makers, providing good coverage for GWAS markers of common and low frequency, biological function and human disease (including Alzheimer's disease, autoimmune/inflammatory, cancer, cardio-metabolic variants, lung function phenotypes, neurological disorders) in the European and British populations. It also includes rare coding variants, pharmacogenomics markers, copy number regions, HLA, inflammation, and eQTL (**Figure 11**). Samples were analysed in batches of approximately 4700 items, and there were 11 batches for UK BiLEVE Axiom array and 95 batches for the UK Biobank Axiom array (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22000).

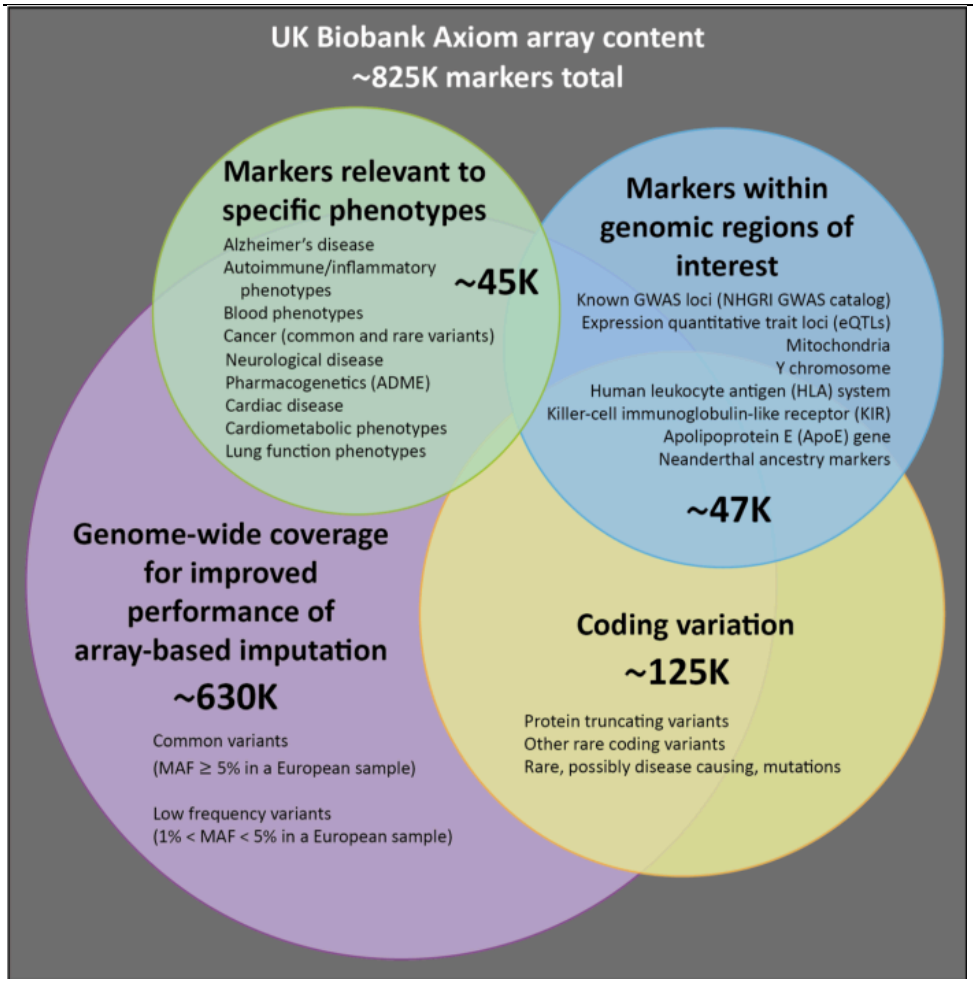


Figure 11. Summary of Affymetrix UK Biobank array content.

A representation of different categories of content on the UK Biobank Axiom array. Numbers indicate the approximate count of markers within each category, ignoring any overlap.

Source: reproduced from Figure 1 of reference 261. This figure was slightly different from the original figure in article because it came from an online available old version of the paper on biorxiv (www.biorxiv.org). This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

The genotyping data of UK Biobank was released in two chunks: the intermediate data release, which contains ~150,000 samples (50,000 UK BiLEVE samples and 100,000 UK Biobank Axiom array), and the full data release of 488, 377 participants.

The UK Biobank imputation was implemented after the array genotyping was done at Affymetrix. Imputation for intermediate data release was done using the programme IMPUTE 2 and was based on a merged reference panel of the Phase 3 of 1000 Genome

Project and the UK10K. For the full data release, the imputation was conducted with both the HRC panel and the merged panel of UK10K and 1000 Genome panel. The intermediate data imputation consisted of 87,696,888 bi-allelic markers in 12,570 haplotypes, and the final data imputation included 92,693,853 autosomal SNPs (261).

4.1.4.2 Quality Control

A number of quality checks were implemented for the processes of sample retrieval, DNA extraction, and genotype callings. Affymetrix also carried out measurements to check the quality for array markers, especially those newly designed markers which were not on Affymetrix array before. Any sample or marker that failed these checks was excluded from the final data release.

Subsequently, genotype data were sent from Affymetrix to the Wellcome Trust Centre for Human Genetics (WTCHG) at the University of Oxford, where, before the full data release, quality control checking marker-based quality and sample quality were conducted (261).

The marker-based quality control included tests on batch effects, plate effects, and departure from Hardy-Weinberg Equilibrium (HWE), sex effects, array effects, and discordance across control replicates (**Table 11**). If a marker failed at least one test in a given batch, the genotype calls in that specific batch were set to missing. If a marker was considered to be unreliable across all batches, the marker was excluded from all the data.

Table 11. UK Biobank marker-based quality control.

Test	Average number of SNPs failed per batch (sd)	Fraction of all genotype calls affected
Affymetrix cluster QC	1109 (699)	0.00140
1. Batch effect	197 (86)	0.000249
2. Plate effect	284 (266)	0.000358
3. Departure from Hardy-	572 (77)	0.000723

Weinberg equilibrium		
4. Sex effect	45 (5)	0.0000569
5. Array effect	5417	0.00683
6. Discordance across controls	622 and 632	0.000796
Total	7704 (721)	0.00971

For all numbered tests a marker was set to missing if the test yielded a p-value < 10^{-12}, except in case of test 6, for which a marker was set to missing if the test yielded < 95% concordance. The array effect test was applied across all batches and only for markers present on both arrays. The discordance test was applied across all batches, but not all markers are present on both arrays. The first value is the number of unique marker on the UK BiLEVE Axiom array that failed this test, and the second is for markers on the UK Biobank Axiom array.

For sample quality control, missingness rate and heterozygosity (derived using a set of ~600k high quality autosomal markers) were used to identify poor quality samples. In addition, with ~15k high quality markers on X and Y chromosome, the sex of each individual was inferred from genotyping data, and confirmed with self-report sex from that participant. UK Biobank did not remove any individuals from their data release for the above reasons but they created tagging variables for individuals who were outliers on the above tests or with probable low sample quality (**Table 12**). However, they excluded 835 samples which were identified as duplicates, ~10 samples which were liable to sample mishandling in the lab, and 33 participants who asked to be withdrawn from UK Biobank.

Table 12. Sample quality control for UK Biobank.

Test	Number of participants
Outliers for heterozygosity/missing rate	968
Putative sex chromosome aneuploidy	652
Sex mismatch	378

4.1.4.3 Ethnic background for UK Biobank participants

During the assessment visit, participants were asked to report their ethnic background. Although most of them self-reported as white, there were a small fraction of other ethnicities (**Table 13**). With a set of 407,219 unrelated, high quality samples and 147,604 high quality markers, the top 40 principal components (PCs) were computed with fastPCA (262), which is an algorithm performing well on very large datasets like UK Biobank. Then PC loadings were computed and all samples were projected onto the PCs, forming a set of PC scores for all samples, which was also included in the data release. **Figure 12** shows results for the first 6 PCs plotted in consecutive pairs. Since the effect of population stratification is an issue which should be considered by epidemiological or genetic studies, UK Biobank also identified 409,703 white British individuals who were all self-reported as British, and who had very similar genetic background based on their PCs.

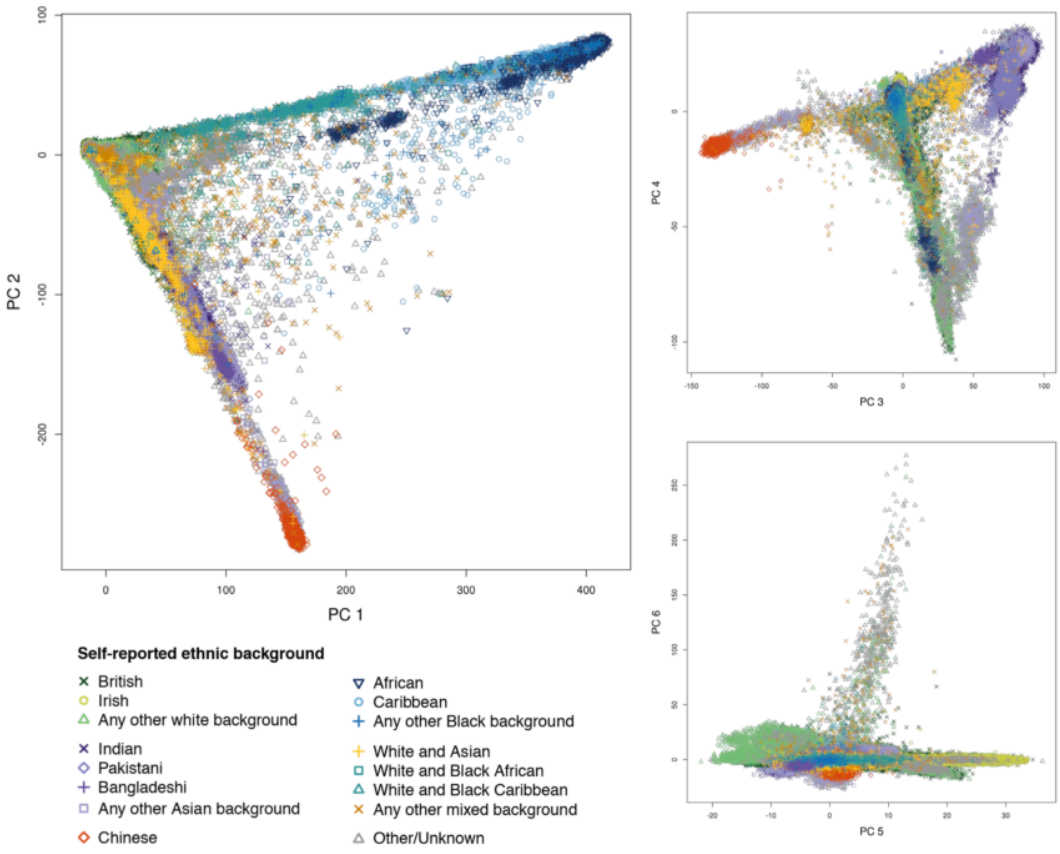


Figure 12. Ancestral diversity in the UK Biobank cohort.

Plots of consecutive pairs of the first six principal components in a PCA of genotype data for UK Biobank participants. Each point represents an individual and is placed according to their principal component scores.

Source: reproduced from Figure 3, reference 261. This figure was slightly different from the original figure in article because it came from an online available old version of the paper on biorxiv (www.biorxiv.org). This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Table 13. Self-report ethnic background of the UK Biobank cohort.

Ethnic group	Self-reported ethnic background	Number (percentage) of UK Biobank participants
White		**472837 (94.24%)**
	British	442710 (88.23%)
	Irish	13215 (2.63%)
	White	571 (0.11%)
	Any other white background	16341 (3.26%)
Asian or Asian British		**9882 (1.97%)**
	Indian	5951 (1.19%)
	Pakistani	1837 (0.37%)
	Any other Asian background	1815 (0.36%)
	Bangladeshi	236 (0.05%)
	Asian or Asian British	43 (0.01%)
Black or Black British		**8066 (1.61%)**
	Caribbean	4520 (0.90%)
	African	3396 (0.68%)
	Any other black background	123 (0.02%)
	Black or Black British	27 (0.01%)
Chinese		**1574 (0.31%)**
	Chinese	1574 (0.31%)
Mixed		**2958 (0.59%)**
	White and Asian	831(0.17%)
	White and Black Caribbean	620 (0.12%)
	White and Black African	425 (0.08%)
	Mixed	49 (0.01%)
	Any other mixed background	1033 (0.21%)
Others/Unknown		**6439 (1.28%)**
	Do not know	217 (0.04%)
	Other ethnic group	4560 (0.91%))

	Prefer not to answer	1662 (0.33%)
Total		**501756 (100%)**

4.1.4.4 Internal relatedness

Relatedness between individuals could confound the results of epidemiological studies, especially when applying general linear models and this familial structure cannot be accounted for well in analyses. Thus, genome or phenome wide analyses try to ensure that participants are independent (i.e., not related). Participants are not independent in the UK Biobank and indeed, many participants are not aware that a close relative is also part of the cohort. To explore familial relatedness among UK Biobank participants, kinship coefficients were estimated for all pairs of samples to identify related individuals. For all pairs of individuals who were inferred to be 3rd degree or closer, the coefficients were reported by UK Biobank. Relationship classes were identified for each related pair using the kinship coefficient and the fraction of markers for which they share no alleles (263). A total of 147,731 individuals (30.3%) were inferred to be related (3rd degree or closer) to at least one other person in the cohort, and form a total of 107,162 related pairs, with a large proportion (66928, 62.45%) of them being 3rd degree related (**Table 14**).

Table 14. Summary of the related pairs (3rd degree or closer) for UK Biobank.

	Monozygotic twins	Parent-offspring	Full sibling	2nd degree	3rd degree	Total
Number of pairs	179	6276	22666	11113	66928	107162

4.2 Statistical Methods

The analysis process compromises three parts: 1) creation of the vitamin D score; 2) Phenome Wide Association analyses; and 3) Mendelian Randomization analyses. I will present details of the 3 major steps and the applied statistical methods in this section. All statistical analyses were implemented in R 3.3.2. For the PheWAS by phecode system, the R package developed by Carroll et al. (207) was exploited.

4.2.1 Creation of the Vitamin D Score

I created a genetic risk score for 25(OH)D. In the selection of variants, I used results from the most recent and largest GWAS- SUNLIGHT 2018, which was conducted with a total of 79,336 individuals of European ancestry (158). This GWAS identified 6 loci associated with serum 25(OH)D concentration and these were rs3755967 (*GC*), rs12785878 (*NADSYN1/DHCR7*), rs10741657 (*CYP2R1*), rs17216707 (*CYP24A1*), rs10745742 (*AMDHD1*) and rs8018720 (*SEC23A*) and they explained a total of 2.84% of the variance of the serum 25OHD level (158). I created a genetic score by adding the number of effect alleles carried by all 6 SNPs and weighted them based on their effect estimates (derived from SUNLIGHT). Imputation data were used in calculation of the score, and the *risk-score* module in software qctool v2.0 was implemented to generate the score from imputation data directly.

4.2.2 Phenome Wide Association Study Method

Phenotyping and mapping ICD codes to phecodes

The "phecode system" was developed by Denny et al. to facilitate the high throughput phenotyping process of PheWAS (details gave in Chapter II). In their state-of-art PheWAS, EMRs of ICD9 codes were first mapped to phecodes and then analyses were performed for predictors on phecodes. As stated in Chapter II, this curating system decreases the number of phenotype groups to be tested in a PheWAS and makes each group more clinically relevant. However, this version of the "phecode system" worked only with the ICD9 clinical modification, an ICD9 version applied in the USA. Collaborating with Professor Denny Joshua and his group, we updated the system so that it could work with ICD10 as well. In this collaboration, I first sent the ICD10 summary data from UK Biobank to Denny's group, under the consent from UK Biobank. The summary data sent by me were generated by counting the numbers of participants with every unique ICD10 code, which had two columns, i.e. ICD10 code and number of participants. On receiving the ICD10 summary data, ICD10 codes were mapped to phecodes by Denny's group either directly or indirectly. If the description of an ICD10 code matches the description of a phecode, it was then mapped to that phecode directly. Otherwise, ICD10 codes were mapped to phecodes indirectly. They used the Unified Medical Language System (UMLS), which integrates and distributes key terminology, classification and coding standards, either to mapped an ICD10 code to an ICD-9-CM code or mapped the ICD10-code to a Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) code first and then to an ICD-9-CM code. Either way, they finally used the previous ICD-9-CM to phecode mapping to link the ICD10 code to a phecode indirectly (**Figure 13**).

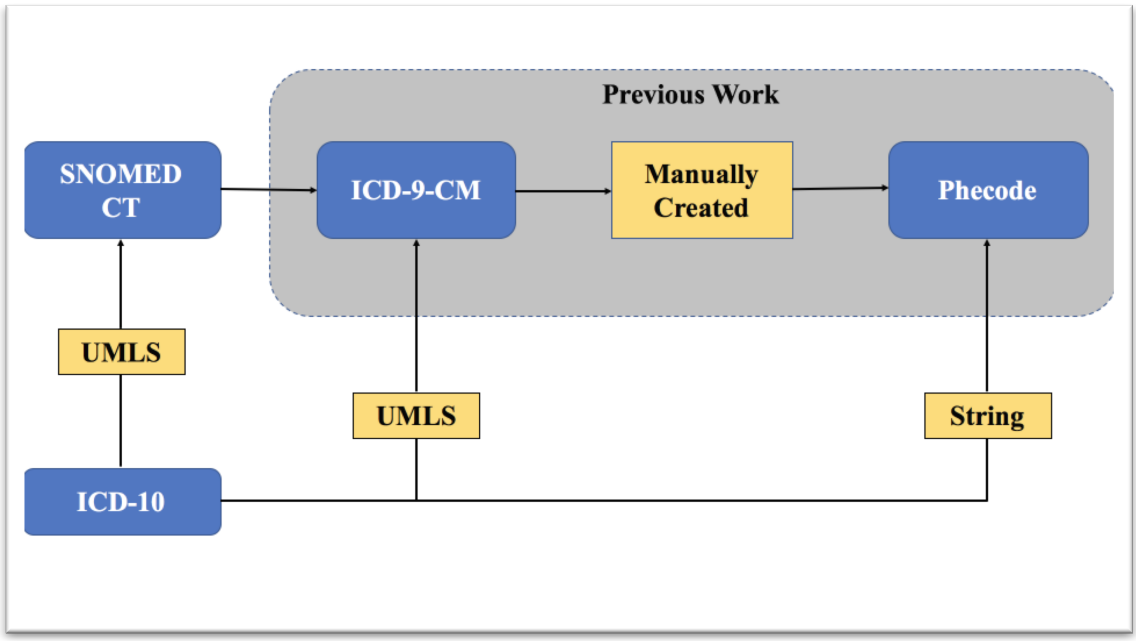


Figure 13. Mapping methods between ICD-10 codes and phecodes.

Abbreviations: SNOMED CT: Systematized Nomenclature of Medical Clinical Terms. UMLS, Unified Medical Language System.

Before my PheWAS analyses, I first pooled the hospital inpatient, cancer registry, and death registry data together. Then for each participant, every distinctive code he/she ever had was aggregated and the number of episodes reporting on that specific code was counted. We then applied the phecode algorithm (199) to define the case/control status of every participant for every phenotype group in the phecode system. For instance, an individual with a record of the ICD10 code M05.2 (rheumatoid vasculitis) was coded as a case for the phecode group 714.1 (rheumatoid arthritis). In addition, this individual was coded as a case for the phecode groups 714 (inflammatory polyarthropathies), 709.7 (Unspecified diffuse connective tissue disease), 716.9 (Arthropathy, not otherwise specified), and 446.3 (Hypersensitivity angiitis). Furthermore, this individual was flagged as an exclusion for phecodes 714-716.0 (other autoimmune arthritis, except for 714, and 714.1, as these are highly related codes and cases for 714.1 and 714 will not be an appropriate controls for these groups), and 696-696.99 (psoriasis, as these codes represent diseases also related to the immune system). Finally, this individual was coded as a control for all other phecode groups

(199). For details of phecode mapping and phecode descriptions, please refer to: https://phewascatalog.org/phecodes.

PheWAS analysis

I first explored the distribution of common confounding variables for health-related outcomes. Subsequently, genotypes of each variant were tested for Hardy-Weinberg equilibrium. In addition, I tested whether genotypes of the 6 SNPs distributed evenly across all UK Biobank assessment centres, in order to explore any population stratification among the included individuals. Chi square tests were implemented, comparing the expected counts to observed counts for the genotypes counts of every genotype in every assessment centre. The associations between the genetic risk score and common confounding factors, including age, gender, body mass index (BMI), physical activity (quantified as outdoor activities), the assessment centre each participant attended (approximation for residing area), average household income, education qualifications and alcohol intake were tested, since they are considered to be associated with many diseases (for the collection and measurement of these variables, please see section 4.1). For continuous variables, univariate linear regression was conducted with the score as the independent and the confounder of interest as the dependent variable. For categorical variables, analysis of variance was conducted.

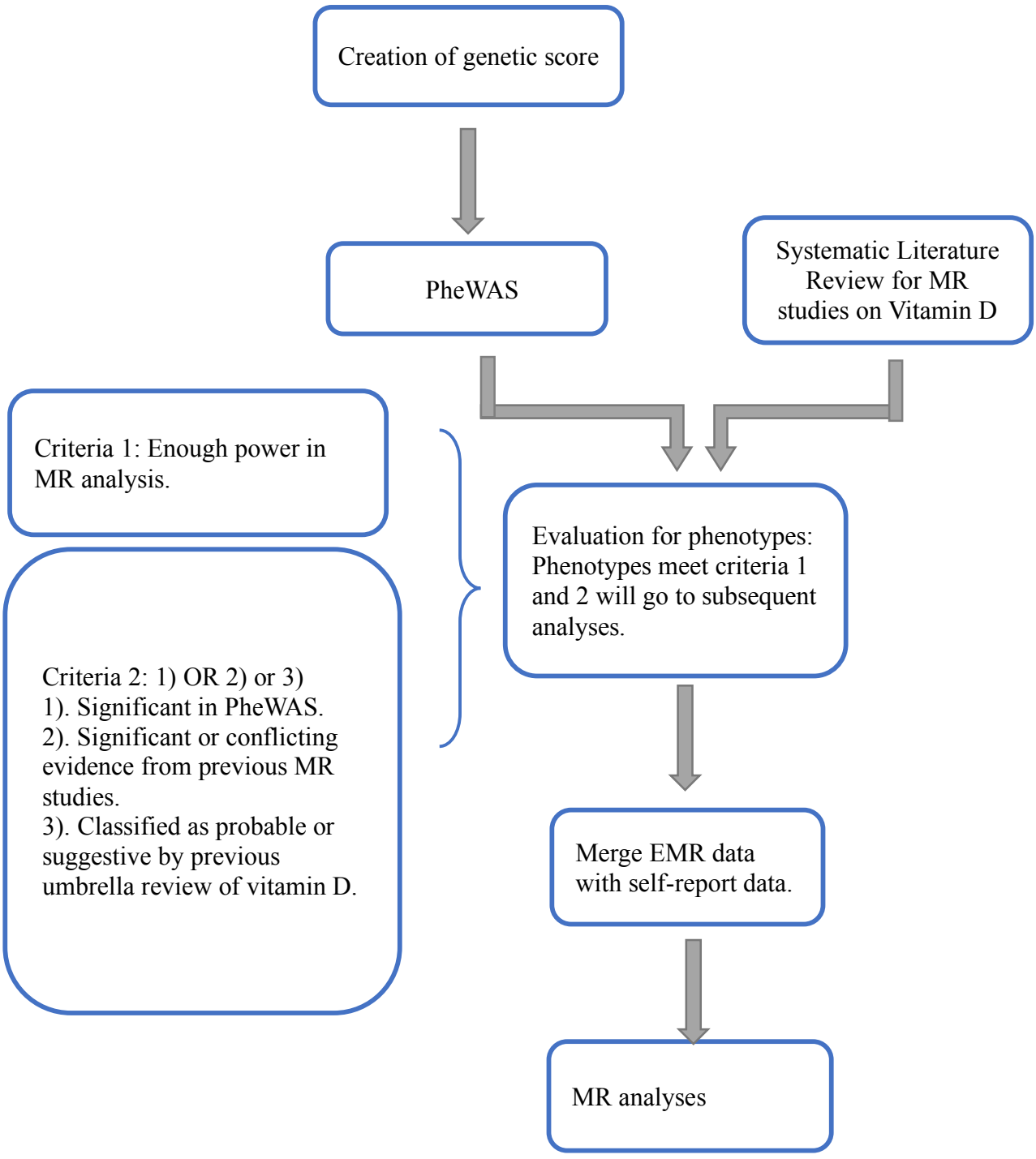
In the PheWAS analysis I only included disease groups with more than 200 cases based on a previous PheWAS simulation study (264). I applied multivariate logistic regression with disease outcome as the dependent variable and the 25(OH)D genetic risk score as the independent adjusted for gender, age, BMI, the UK Biobank assessment centre attended, east and north co-ordinate of home address and the first 5 genetic PCs. I tested a total of 920 disease groups and thus a *P*-value of less than $5.44*10^{-5}$ was regarded as statistically significant based on Bonferroni correction ($0.05/920 = 5.44*10^{-5}$). In addition, each of the 6 SNPs (i.e., rs3755967 (*GC*), rs12785878 (*NADSYN1/DHCR7*), rs10741657 (*CYP2R1*), rs17216707 (*CYP24A1*), rs10745742 (*AMDHD1*) and rs8018720 (*SEC23A*)) were tested on their associations with all phenotype groups of more than 200 cases individually, adjusting for the same variables.

4.2.3 Mendelian Randomization Methods

Outcome selection

Mendelian Randomization studies are of low power compared with traditional observational studies, partly caused by the small variance of biomarker explained by the genetic instrumental variable (265). Thus, only phenotypes with sufficient power in a Mendelian Randomization context need to be tested in my MR analyses. In my study, I considered those phenotypes with more than 80% power for detecting a true OR of 1.2 or greater in a Mendelian Randomization study, assuming an explained variance of 3%, case/control ratio of 1/5 or smaller (which should always be true with a large sample like UK Biobank), at an alpha level of 0.05, as of sufficient statistical power. I conducted the power calculation for binary outcomes based on an equation published by Burgess S. (265). Phenotypes with enough power which survived Bonferroni correction for the PheWAS between the 25(OH)D score and phenotypes (the results from single SNP PheWAS are not eligible for MR analyses, since I am trying the study the causal association between 25(OH)D as a biomarker and health outcomes) or showed probable or suggestive level evidence on association with vitamin D from the previous umbrella review on observational studies and RCTs (25) or were statistically significantly associated with vitamin D in existing MR studies, or those with conflicting results from existing MR studies were selected to be studied in my MR analyses (**Figure 14**).

Figure 14. Outcome selection for Mendelian Randomization analysis



Merge EMR data and self-reported data

For outcomes eligible for MR analyses, I merged UK Biobank EMR data with self-reported medical conditions which were reported by each participant in the baseline questionnaire and the subsequent verbal interview with research nurse (details gave in section 4.1). By merging the self-reported data in, I aimed to identify cases which were missed by EMR data and thus falsely classified as controls in previous PheWAS, and to increase the statistical power further.

Mendelian Randomization analyses

I then ran MR analyses using multiple methods, including: a) two-stage MR, b) inverse variance weighted (IVW) MR and c) Egger's regression MR.

To test the robustness of my risk score as IV, I constructed the identical weighted sum score of 6 SNPs in the Study of Colorectal Cancer in Scotland (SOCCS) (266) and then tested the variance of 25(OH)D explained by this score and the corresponding *F* statistics. Furthermore, the standard deviation of the log transformed 25(OH)D level was also derived from SOCCS, which was used in subsequent calculation of OR values using a metric of per standard deviation change of log transformed 25(OH)D level as determined by genetic instruments.

The score I used in the PheWAS is a weighted sum score whose weights came from SUNLIGHT 2018. In my study, the SUNLIGHT 2018 study represented the first stage of two-stage MR, testing the association between SNPs and the selected biomarker, vitamin D. The score created was just the predicted vitamin D level, as determined by genetic variants. I then ran the second stage of the two-stage MR, a logistic regression with the outcome variable (merged EMR and self-reported data) as dependent variable and the 25(OH)D score as independent variable, adjusting for gender, participants' age while attending baseline measurements, BMI, the UK Biobank assessment centre attended, east and north co-ordinate of home address and the first 5 ancestral PCs. The exponential of the coefficient of the regression is the MR estimate of the causal effect of 25(OH)D on outcome.

I then conducted inverse variance weighted (IVW) MR and Egger's MR to check the consistency of my results between different MR methods and account for any unbalanced pleiotropy of individual SNPs. Causal effect estimates from separate SNPs as instrumental variables can be pooled together using the IVW method. In brief, individual effect estimates are first calculated by the ratio method for every SNP, where the coefficient from the regression of outcome on individual SNP is divided by the effect estimates of SNP on biomarker exposure. Then, individual effect estimates are pooled together by a random-effects meta-analysis (267). The overall causal effect from multiple IVs estimated by IVW equals the coefficient from a weighted regression of IV-outcome association on IV-exposure association with the intercept constrained at zero. However, this method assumes that IVs are not associated with outcomes conditional on exposure and confounders (i.e., all the effect of IVs on outcome are conducted via the exposure), which is also called "exclusion restriction". Pleiotropic effects of IVs would cause potential violation of this assumption. Egger's MR employs a less strict version of exclusion restriction, the InSIDE, where the correlation between the genetic associations with the exposure and the direct effects of the genetic variants on the outcome is zero (268). By viewing a MR with multiple IVs as analogous to a meta-analysis, the bias caused by pleiotropy is analogous to small study bias in meta-analysis. Under the InSIDE assumption, applying a regression method for all the individual IVs without constraining the intercept to be zero, Egger's MR tests whether there is unbalanced pleiotropy (rejection of the null hypothesis of the intercept value equal to zero implies unbalanced pleiotropy, where there are pleiotropic effects with IVs and they do not cancel out). It gives a summary effect estimates closer to truth compared with IVW, especially under situations of unbalanced pleiotropy (268).

Finally, to test the robustness of the results from my MR analyses, two sensitivity analyses were implemented for outcomes in MR analyses. These include leave-one-out MR analyses and MR analyses excluding variants liable to population stratification. In the leave-one-out analyses, I drop one SNP each time, and conduct IVW MR analyses involving the other 5 SNPs, which is used to explore the possibility that the overall MR result of 6 SNPs is dominated by one single SNP. Then, in the other sensitivity analyses, I exclude variants which show evidence of population

stratification and conduct IVW MR analyses involving all other SNPs to control the influence of population stratification on my results.

Chapter V: Results

5.1 Association between genetic variants and phenotypes

5.1.1 Descriptive statistics

A total of 339,256 unrelated British white individuals were included in the analysis of this study. Of the included individuals, 182110 (53.86%) were females. The mean age was 56.89 years (standard deviation = 7.99), and the mean BMI was 27.40 kg/m^2 (standard deviation = 4.76) (**Table 15**).

Table 15. Demographic characteristics of the UK Biobank participants.

Variable	Value
Demographic characteristics (n=339,256)	
Female	182,110 (53.68%)
Age	56.89 (7.99) years
BMI	27.40 (4.76) kg/m^2

Abbreviations: BMI, body mass index, which is derived by weight (in kg) divided square of height (in metres).

Notes: Continuous variables are presented as mean (standard deviation), while categorical variables are presents as N (%).

The participants were of mean standing height of 168.80 cm (standard deviation = 9.24); mean weight of 78.31 kg (standard deviation = 15.88). As an important confounding factor for vitamin D level, outdoor activity was also considered in my analyses. Mean time spend outdoors in summer by participant was 3.17 hours/day (standard deviation = 3.58). Mean time spend outdoors in winter by participants was 0.14 hour/day (standard deviation = 4.71). Other confounding factors I explored included household income before tax, qualifications and alcohol intake frequency. For details, please see **Table 16**.

Household income before tax is reported as a categorical variable with 7 levels, which were “less than £18,000”, “£18,000 to £30,999”, “£31,000 to £51,999”, “£52,000 to £100,000”, “greater than £100,000”, “Do not know”, “Prefer not to answer”. A total of 1108 participants did not report their income before tax. Among the available data,

there were 63806 (18.81%), 75135 (22.15%), 77268 (22.78%), 60478 (17.83%), 15678 (4.62%), 13027 (3.84%), 32756 (9.66%) for each of the above category. Combined with "Do not know" and "Prefer not to answer", I did not have valid data for a total of 46891 (13.82%) participants.

Qualifications were also reported as a categorical variable from the baseline touchscreen questionnaire. The eight levels were "College or University degree", "A levels/AS levels or equivalent", "O levels/GCSEs or equivalent", "CSEs or equivalent", "NVQ or HND or HNC or equivalent", "Other professional qualifications (eg: nursing, teaching)", "None of the above", "Prefer not to answer". There were 319 participants who did not select any category, and 107150 (31.58%), 38467 (11.34%), 74674 (22.01%), 18231 (5.37%), 22411 (6.61%), 17418 (5.13%), 57809 (17.04%) and 2777 (0.82%) participants who selected each of the above level of education. I did not have valid information for 60905 (17.95%) participants.

Alcohol intake frequency was reported as a categorical variable of 7 levels, which were "daily or almost daily", "3 or 4 times a week", "1 or 2 times a week", "1 to 3 time a month", "special occasions only", "never" and "prefer not to answer". There were 72786 (21.45%), 82022 (24.18%), 89054 (26.25%), 37492 (11.05%), 35601 (10.49%), 22065 (6.50%) and 236 (0.07%) participants selecting each of the level above. All participants reported data for this variable. However, due to the options of "never" and "prefer not to answer", I still did not have valid data for a total of 22301 (6.57%) participants.

Table 16. Characteristics for confounding factors of the UK Biobank participants.

Variable Name	**Feature**	**N (%)**
Continuous	Mean (S.D.)	Number of missingness
Standing height	168.80 (9.24) cm	730 (0.22%)
Weight	78.31 (15.88) kg	964 (0.28%)
Time spend outdoors in summer	3.17 (3.58) hours/day	319 (0.09%)
Time spend outdoors in winter	0.14 (4.71) hours/day	319 (0.09%)

Categorical	Levels	Number of participants
Household income before tax	Less than £18,000	63,806 (18.81%)
	£18,000 to £30,999	75,135 (22.15%)
	£31,000 to £51,999	77,268 (22.78%)
	£52,000 to £100,000	60,478 (17.83%)
	Greater than £100,000	15,678 (4.62%)
	Do not know	13,027 (3.84%)
	Prefer not to answer	32,756 (9.66%)
	Missingness	1108 (0.33%)
Qualifications	College or University degree	107,150 (31.58%)
	A levels/AS levels or equivalent	38,467 (11.34%)
	O levels/GCSEs or equivalent	74,674 (22.01%)
	CSEs or equivalent	18,231 (5.37%)
	NVQ or HND or HNC or equivalent	22,411 (6.61%)
	Other professional qualifications (eg: nursing, teaching)	17,418 (5.13%)
	None of the above	57,809 (17.04%)
	Prefer not to answer	2,777 (0.82%)
	Missingness	319 (0.09%)
Alcohol intake frequency	Daily or almost daily	72,786 (21.45%)
	3 or 4 times a week	82,022 (24.18%)
	1 or 2 times a week	89,054 (26.25%)
	1 to 3 times a month	37,492 (11.05%)
	Special occasions only	35,601 (10.49%)
	Never	22,065 (6.50%)
	Prefer not to answer	236 (0.07%)
	Missingness	0

Baseline measurements for these participants were taken in 22 UK Biobank assessment centres. For the number participants attending each assessment centre, please see **Table 17**. The range of the counts spread from 319 to 31348. Stockport was an assessment centre for the pilot part of the study only, thus only 319 (0.09%) participants were from Stockport. Leeds recruited the largest number of participants, which was 31348 (9.24%).

Table 17. Number of participants attending each assessment centre.

Assessment Centre	N (%)
Barts	5,908 (1.74%)
Birmingham	16,056 (4.73%)
Bristol	30,830 (9.09%)
Bury	20,383 (6.01%)
Cardiff	12,781 (3.77%)
Croydon	15,076 (4.44%)
Edinburgh	12,355 (3.64%)
Glasgow	12,693 (3.74%)
Hounslow	14,763 (4.35%)
Leeds	31,348 (9.24%)
Liverpool	22,685 (6.69%)
Manchester	9078 (2.68%)
Middlesborough	15,295 (4.51%)
Newcastle	26,285 (7.75%)
Nottingham	24,411 (7.20%)
Oxford	9,922 (2.92%)
Reading	21,249 (6.26%)
Sheffield	21,885 (6.45%)
Stockport(pilot)	319 (0.09%)
Stoke	13,824 (4.07%)
Swansea	1,620 (0.48%)
Wrexham	490 (0.14%)

5.1.2 Genotypes of vitamin D related genetic variants

According to the most recent GWAS on 25(OH)D level at the time of the analysis, six SNPs were statistically significantly associated with 25(OH)D level at a GWAS *P*-value threshold ($P < 10^{-8}$) (i.e., rs3755967 (*GC*), rs10741657 (*CYP2R1*), rs12785878 (*DHCR7*), rs10745742 (*AMDHD1*), rs8018720 (*SEC23A*), rs17216707(*CYP24A1*)). I extracted the genotypes of these 6 SNPs from the imputation data, and genotype distributions of them were described in **Table 18**. All 6 SNPs were in Hardy-Weinberg disequilibrium.

Table 18. Genotype counts of the six vitamin D related SNPs.

Genotype counts	N (%)
rs3755967 polymorphism (n=338,753)	
CC	169,710 (50.10%)
CT	140,206 (41.39%)
TT	28,837 (8.51%)
Hardy-Weinberg *P*-value	0.52
rs10741657 polymorphism (n=339,256)	
AA	55,617 (16.39%)
AG	163,064 (48.07%)
GG	120,575 (35.54%)
Hardy-Weinberg *P*-value	0.83
rs12785878 polymorphism (n=339,256)	
TT	211,627 (62.38%)
TG	112,585 (33.19%)
GG	15,044 (4.43%)
Hardy-Weinberg *P*-value	0.35
rs10745742 polymorphism (n=336,987)	
TT	47,797 (14.18%)
TC	158,392 (47.00%)
CC	130,798 (38.82%)
Hardy-Weinberg *P*-value	0.29
rs8018720 polymorphism (n=339,256)	
GG	10,666 (3.14%)
GC	98,435 (29.02%)
CC	230,155 (67.84%)
Hardy-Weinberg *P*-value	0.80
rs17216707 polymorphism (n=324,016)	
TT	216,735 (66.89%)
TC	96,403 (29.75%)
CC	10878 (3.36%)
Hardy-Weinberg *P*-value	0.78

The frequency for the T allele of rs3755967 (*GC*) varied from 0.285 (Stockport) to 0.310 (Wrexham) (**Table 19**). The *P* value for the χ^2 test for rs3755967 was 0.105, suggesting that this variant was distributed evenly across UK Biobank assessment centres.

Table 19. Genotype distribution for rs3755967 across UK Biobank assessment centres.

Assessment centre	Genotype	Value
Barts		
	CC	2915
	CT	2475
	TT	506
	Frequency of T allele	0.296
Birmingham		
	CC	8039
	CT	6676
	TT	1318
	Frequency of T allele	0.290
Bristol		
	CC	15366
	CT	12751
	TT	2679
	Frequency of T allele	0.294
Bury		
	CC	10216
	CT	8461
	TT	1664
	Frequency of T allele	0.290
Cardiff		
	CC	6285
	CT	5406
	TT	1074
	Frequency of T allele	0.296
Croydon		
	CC	7454
	CT	6299
	TT	1300
	Frequency of T allele	0.296
Edinburgh		
	CC	6251
	CT	5059
	TT	1024
	Frequency of T allele	0.288
Glasgow		

	CC	6395
	CT	5197
	TT	1074
	Frequency of T allele	0.290
Hounslow		
	CC	7301
	CT	6121
	TT	1319
	Frequency of T allele	0.297
Leeds		
	CC	15849
	CT	12873
	TT	2587
	Frequency of T allele	0.288
Liverpool		
	CC	11400
	CT	9325
	TT	1929
	Frequency of T allele	0.291
Manchester		
	CC	4602
	CT	3686
	TT	771
	Frequency of T allele	0.289
Middlesborough		
	CC	7677
	CT	6287
	TT	1308
	Frequency of T allele	0.291
Newcastle		
	CC	13330
	CT	10771
	TT	2148
	Frequency of T allele	0.287
Nottingham		
	CC	12191
	CT	10072
	TT	2114

	Frequency of T allele	0.293
Oxford		
	CC	4853
	CT	4153
	TT	899
	Frequency of T allele	0.300
Reading		
	CC	10646
	CT	8783
	TT	1791
	Frequency of T allele	0.291
Sheffield		
	CC	10834
	CT	9077
	TT	1941
	Frequency of T allele	0.297
Stockport (pilot)		
	CC	165
	CT	123
	TT	29
	Frequency of T allele	0.285
Stoke		
	CC	6904
	CT	5722
	TT	1182
	Frequency of T allele	0.293
Swansea		
	CC	798
	CT	691
	TT	127
	Frequency of T allele	0.292
Wrexham		
	CC	239
	CT	198
	TT	53
	Frequency of T allele	0.310
Overall χ^2 test	$P = 0.105$	

The frequency for the G allele of rs10741657 (*CYP2R1*) varied from 0.562 (Wrexham) to 0.606 (Barts) (**Table 20**). The *P* value for the χ^2 test for rs10741657 was 0.003, implying that this variant was distributed unevenly across UK Biobank assessment centres. I then re-coded assessment centres into two categories (Scotland or England/Wales), in order to test the distribution of this SNP across latitude. A total of 25048 participants (7.38%) were recruited from Edinburgh or Glasgow. However, the *P* value for this χ^2 test was 0.878. Thus, there was no difference between genotype of rs10741657 in Scotland and genotype of that in England and Wale, which is of lower latitude. In χ^2 test of participant from Wrexham (n = 490, 0.14%) vs all others, the *P* value was 0.0613. In χ^2 test of participant from Wrexham or Cardiff (n = 13271, 3.91%) vs all other the *P* value was 0.001. Participants from Wrexham and Cardiff were of lowest level of G allele frequency of rs10741657, and it is significantly different from all other assessment centres.

Table 20. Genotype distribution for rs10741657 across UK Biobank assessment centres.

Assessment centre	Genotype	Value
Barts		
	AA	942
	AG	2776
	GG	2190
	Frequency of G allele	0.606
Birmingham		
	AA	2582
	AG	7749
	GG	5725
	Frequency of G allele	0.598
Bristol		
	AA	4938
	AG	14635
	GG	11257
	Frequency of G allele	0.602
Bury		
	AA	3354
	AG	9840
	GG	7189
	Frequency of G allele	0.594
Cardiff		
	AA	2146
	AG	6277
	GG	4358
	Frequency of G allele	0.587
Croydon		
	AA	2440
	AG	7280
	GG	5356
	Frequency of G allele	0.597
Edinburgh		
	AA	2028
	AG	5948
	GG	4379
	Frequency of G allele	0.595

Glasgow		
	AA	2096
	AG	6109
	GG	4488
	Frequency of G allele	0.594
Hounslow		
	AA	2394
	AG	7119
	GG	5250
	Frequency of G allele	0.597
Leeds		
	AA	5106
	AG	15016
	GG	11226
	Frequency of G allele	0.598
Liverpool		
	AA	3836
	AG	10925
	GG	7924
	Frequency of G allele	0.590
Manchester		
	AA	1475
	AG	4377
	GG	3226
	Frequency of G allele	0.596
Middlesborough		
	AA	2516
	AG	7400
	GG	5379
	Frequency of G allele	0.594
Newcastle		
	AA	4473
	AG	12658
	GG	9154
	Frequency of G allele	0.589
Nottingham		
	AA	3933
	AG	11741

	GG	8737
	Frequency of G allele	0.598
Oxford		
	AA	1692
	AG	4737
	GG	3493
	Frequency of G allele	0.591
Reading		
	AA	3457
	AG	10132
	GG	7660
	Frequency of G allele	0.599
Sheffield		
	AA	3550
	AG	10484
	GG	7851
	Frequency of G allele	0.598
Stockport (pilot)		
	AA	52
	AG	154
	GG	113
	Frequency of G allele	0.596
Stoke		
	AA	2239
	AG	6730
	GG	4855
	Frequency of G allele	0.595
Swansea		
	AA	269
	AG	746
	GG	605
	Frequency of G allele	0.604
Wrexham		
	AA	99
	AG	231
	GG	160
	Frequency of G allele	0.562
Overall χ^2 test	$P = 0.003$	

The frequency for the G allele of rs12785878 (*DHCR7/NADSYN1*) varied from 0.190 (Glasgow) to 0.223 (Oxford) (**Table 21**). From the table, I could find a trend that participants from centres in the south had a higher G allele frequency compared with participants from middle England (e.g., Leeds, Manchester), while participants from middle England were of higher G allele frequency compared with participants from Scotland (i.e., Edinburgh and Glasgow). The *P* value for the χ^2 test for rs12785878 was $< 2.2*10^{-16}$, implying that this variant was distributed unevenly across UK Biobank assessment centres. I re-coded assessment centres into two categories (Scotland or England/Wales), in order to test the distribution of this SNP across region. A total of 25048 participants (7.38%) were recruited from Edinburgh or Glasgow. The *P* value for χ^2 test was lower than $2.2*10^{-16}$. Thus, the association between rs12785878 genotype and assessment centre was caused by the latitudinal distribution of rs12785878.

Table 21. Genotype distribution for rs12785878 across UK Biobank assessment centres.

Assessment centre	Genotype	Value
Barts		
	TT	3554
	TG	2077
	GG	277
	Frequency of G allele	0.223
Birmingham		
	TT	9909
	TG	5360
	GG	787
	Frequency of G allele	0.216
Bristol		
	TT	18872
	TG	10483
	GG	1475
	Frequency of G allele	0.218
Bury		
	TT	13033
	TG	6541
	GG	809
	Frequency of G allele	0.200
Cardiff		
	TT	7997
	TG	4232
	GG	552
	Frequency of G allele	0.209
Croydon		
	TT	9139
	TG	5155
	GG	782
	Frequency of G allele	0.223
Edinburgh		
	TT	8083
	TG	3816
	GG	456
	Frequency of G allele	0.191

Glasgow		
	TT	8322
	TG	3921
	GG	450
	Frequency of G allele	0.190
Hounslow		
	TT	9057
	TG	4974
	GG	732
	Frequency of G allele	0.218
Leeds		
	TT	19715
	TG	10317
	GG	1316
	Frequency of G allele	0.207
Liverpool		
	TT	14155
	TG	7565
	GG	965
	Frequency of G allele	0.209
Manchester		
	TT	5755
	TG	2934
	GG	389
	Frequency of G allele	0.204
Middlesborough		
	TT	9503
	TG	5088
	GG	704
	Frequency of G allele	0.212
Newcastle		
	TT	16731
	TG	8473
	GG	1081
	Frequency of G allele	0.202
Nottingham		
	TT	15050
	TG	8264

	GG	1097
	Frequency of G allele	0.214
Oxford		
	TT	5976
	TG	3462
	GG	484
	Frequency of G allele	0.223
Reading		
	TT	12836
	TG	7360
	GG	1053
	Frequency of G allele	0.223
Sheffield		
	TT	13789
	TG	7178
	GG	918
	Frequency of G allele	0.206
Stockport (pilot)		
	TT	209
	TG	98
	GG	12
	Frequency of G allele	0.191
Stoke		
	TT	8617
	TG	4589
	GG	618
	Frequency of G allele	0.211
Swansea		
	TT	1007
	TG	541
	GG	72
	Frequency of G allele	0.211
Wrexham		
	TT	318
	TG	157
	GG	15
	Frequency of G allele	0.191
Overall χ^2 test	$P < 2.2 * 10^{-16}$	

The frequency for the C allele of rs10745742 (*AMDHD1*) varied from 0.610 (Leeds) to 0.661 (Swansea) (**Table 22**). Overall, the *P* value for the χ^2 test was < $2.2*10^{-16}$, implying that rs10745742 was distributed unevenly across UK Biobank assessment centres. Then χ^2 test was implemented for participants from Scotland vs all others, and the *P* value was $1.636*10^{-8}$. However, participants from Edinburgh (0.628) had a relatively low frequency of C allele, which was similar to those from the southern centres, compared with those from Glasgow (0.661). This could not be attributed to simply a latitude effect. Participants from Cardiff, Glasgow, Wrexham and Swansea were of higher C allele frequency compared with other centres. Thus, I tested the relevant difference in another χ^2 test, and found a *P* value of lower than $2.2*10^{-16}$.

Table 22. Genotype distribution for rs10745742 across UK Biobank assessment centres.

Assessment centre	Genotype	Value
Barts		
	TT	824
	TC	2723
	CC	2323
	Frequency of C allele	0.628
Birmingham		
	TT	2257
	TC	7581
	CC	6112
	Frequency of C allele	0.621
Bristol		
	TT	4333
	TC	14357
	CC	11943
	Frequency of C allele	0.624
Bury		
	TT	2991
	TC	9578
	CC	7694
	Frequency of C allele	0.616
Cardiff		
	TT	1684
	TC	5802
	CC	5212
	Frequency of C allele	0.639
Croydon		
	TT	2121
	TC	7075
	CC	5763
	Frequency of C allele	0.622
Edinburgh		
	TT	1666
	TC	5791
	CC	4810
	Frequency of C allele	0.628

Glasgow		
	TT	1639
	TC	5729
	CC	5245
	Frequency of C allele	0.643
Hounslow		
	TT	2060
	TC	6900
	CC	5699
	Frequency of C allele	0.624
Leeds		
	TT	4692
	TC	14900
	CC	11576
	Frequency of C allele	0.610
Liverpool		
	TT	3022
	TC	10569
	CC	8943
	Frequency of C allele	0.631
Manchester		
	TT	1269
	TC	4188
	CC	3561
	Frequency of C allele	0.627
Middlesborough		
	TT	2191
	TC	6996
	CC	6010
	Frequency of C allele	0.626
Newcastle		
	TT	3624
	TC	12165
	CC	10315
	Frequency of C allele	0.628
Nottingham		
	TT	3477
	TC	11499

	CC	9283
	Frequency of C allele	0.620
Oxford		
	TT	1367
	TC	4693
	CC	3796
	Frequency of C allele	0.623
Reading		
	TT	3044
	TC	9901
	CC	8160
	Frequency of C allele	0.621
Sheffield		
	TT	3243
	TC	10382
	CC	8094
	Frequency of C allele	0.612
Stockport (pilot)		
	TT	47
	TC	151
	CC	118
	Frequency of C allele	0.612
Stoke		
	TT	2022
	TC	6437
	CC	5247
	Frequency of C allele	0.618
Swansea		
	TT	166
	TC	761
	CC	684
	Frequency of C allele	0.661
Wrexham		
	TT	58
	TC	214
	CC	210
	Frequency of C allele	0.658
Overall χ^2 test	$P < 2.2*10\text{-}16$	

The frequency for the C allele of rs8018720 varied from 0.817 (Wrexham) to 0.832 (Stockport) (**Table 23**). The *P* value for the χ^2 test for rs8018720 was 0.204, supporting that this variant distributed evenly across UK Biobank assessment centres.

Table 23. Genotype distribution for rs8018720 across UK Biobank assessment centres.

Assessment centre	Genotype	Value
Barts		
	GG	178
	GC	1712
	CC	4018
	Frequency of C allele	0.825
Birmingham		
	GG	457
	GC	4697
	CC	10902
	Frequency of C allele	0.825
Bristol		
	GG	996
	GC	8903
	CC	20931
	Frequency of C allele	0.823
Bury		
	GG	619
	GC	6007
	CC	13757
	Frequency of C allele	0.822
Cardiff		
	GG	410
	GC	3735
	CC	8636
	Frequency of C allele	0.822
Croydon		
	GG	468
	GC	4316
	CC	10292

	Frequency of C allele	0.826
Edinburgh		
	GG	421
	GC	3648
	CC	8286
	Frequency of C allele	0.818
Glasgow		
	GG	403
	GC	3766
	CC	8524
	Frequency of C allele	0.820
Hounslow		
	GG	457
	GC	4186
	CC	10120
	Frequency of C allele	0.827
Leeds		
	GG	1017
	GC	9056
	CC	21275
	Frequency of C allele	0.823
Liverpool		
	GG	739
	GC	6619
	CC	15327
	Frequency of C allele	0.822
Manchester		
	GG	321
	GC	2601
	CC	6156
	Frequency of C allele	0.821
Middlesborough		
	GG	492
	GC	4383
	CC	10420
	Frequency of C allele	0.825
Newcastle		
	GG	806

	GC	7794
	CC	17685
	Frequency of C allele	0.821
Nottingham		
	GG	752
	GC	7014
	CC	16645
	Frequency of C allele	0.826
Oxford		
	GG	286
	GC	2922
	CC	6714
	Frequency of C allele	0.824
Reading		
	GG	661
	GC	6132
	CC	14456
	Frequency of C allele	0.825
Sheffield		
	GG	714
	GC	6267
	CC	14904
	Frequency of C allele	0.824
Stockport (pilot)		
	GG	6
	GC	95
	CC	218
	Frequency of C allele	0.832
Stoke		
	GG	404
	GC	3958
	CC	9462
	Frequency of C allele	0.828
Swansea		
	GG	45
	GC	473
	CC	1102
	Frequency of C allele	0.826

Wrexham		
	GG	14
	GC	151
	CC	325
	Frequency of C allele	0.817
Overall χ^2 test	$P = 0.204$	

The frequency for the C allele of rs17216707 varied from 0.176 (Swansea) to 0.201 (Wrexham) (**Table 24**). The *P* value for the χ^2 test for rs17216707 was 0.920, suggesting that this variant distributed evenly across UK Biobank assessment centres.

Table 24. Genotype distribution for rs17216707 across UK Biobank assessment centres.

Assessment centre	Genotype	Value
Barts		
	TT	3749
	TC	1683
	CC	185
	Frequency of C allele	0.183
Birmingham		
	TT	10242
	TC	4551
	CC	538
	Frequency of C allele	0.184
Bristol		
	TT	19756
	TC	8719
	CC	1002
	Frequency of C allele	0.182
Bury		
	TT	13120
	TC	5773
	CC	642
	Frequency of C allele	0.181
Cardiff		
	TT	8242
	TC	3559
	CC	409
	Frequency of C allele	0.179
Croydon		
	TT	9540
	TC	4298
	CC	518

	Frequency of C allele	0.186
Edinburgh		
	TT	7975
	TC	3454
	CC	398
	Frequency of C allele	0.180
Glasgow		
	TT	8212
	TC	3575
	CC	409
	Frequency of C allele	0.180
Hounslow		
	TT	9376
	TC	4177
	CC	503
	Frequency of C allele	0.184
Leeds		
	TT	20076
	TC	8923
	CC	987
	Frequency of C allele	0.182
Liverpool		
	TT	14549
	TC	6371
	CC	727
	Frequency of C allele	0.181
Manchester		
	TT	5798
	TC	2589
	CC	272
	Frequency of C allele	0.181
Middlesborough		
	TT	9739
	TC	4371
	CC	490
	Frequency of C allele	0.183
Newcastle		
	TT	16853

	TC	7485
	CC	835
	Frequency of C allele	0.182
Nottingham		
	TT	15547
	TC	6941
	CC	786
	Frequency of C allele	0.183
Oxford		
	TT	6267
	TC	2885
	CC	316
	Frequency of C allele	0.186
Reading		
	TT	13484
	TC	6092
	CC	694
	Frequency of C allele	0.185
Sheffield		
	TT	13871
	TC	6295
	CC	656
	Frequency of C allele	0.183
Stockport (pilot)		
	TT	195
	TC	93
	CC	10
	Frequency of C allele	0.190
Stoke		
	TT	8795
	TC	3980
	CC	430
	Frequency of C allele	0.183
Swansea		
	TT	1050
	TC	447
	CC	49
	Frequency of C allele	0.176

Wrexham		
	TT	299
	TC	142
	CC	22
	Frequency of C allele	0.201
Overall χ^2 test	$P = 0.920$	

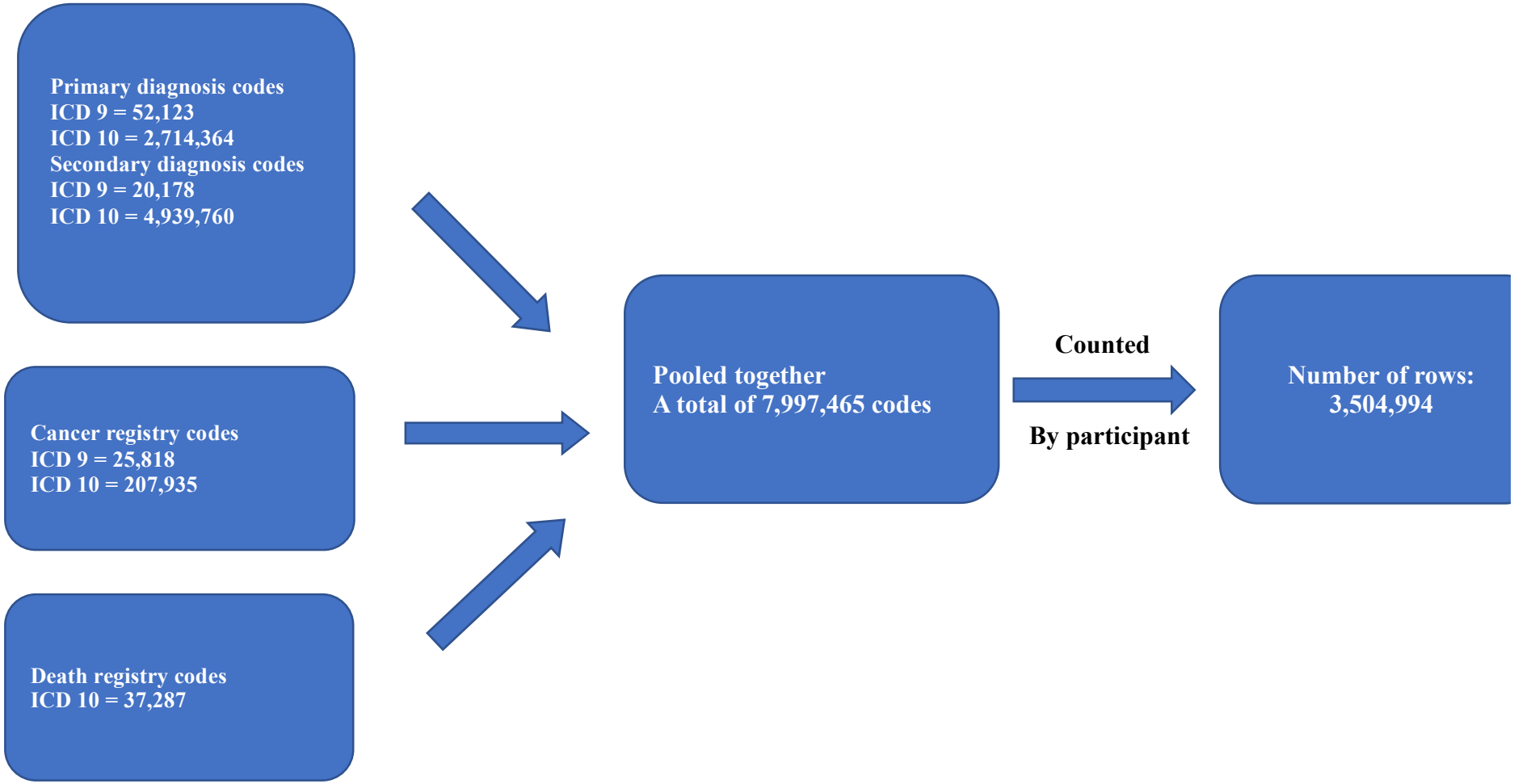
5.1.3 Downloading and manipulation of EMRs

I downloaded the Hospital Episode Statistics in December 2016. The inpatient data I downloaded compromised 2,779,598 unique records of hospital inpatient data corresponding to 395,978 unique individuals (2,714,364 records had an ICD10 diagnosis code, 52,123 had an ICD9 diagnosis code and 13,111 records did not have a diagnosis code). The inpatient data were organized on an episode basis. Each hospitalized episode for a participant generated a unique line in the database. Thus, an individual could have several lines of records in the inpatient database based on the number of episodes he/she ever had. Not every individual had records in the inpatient database.

Cancer registry and death registry were released as separate databases from the inpatient data and I downloaded them as well. In the cancer registry data, there were 233,753 records corresponding to 79,066 unique participants (207,935 records had an ICD10 diagnosis code, and 25,818 had an ICD9 diagnosis code). In addition, 14,417 deaths were recorded in the death registry data.

The phecode system for curating ICD-based phenotyping (ICD10/ICD9 to phecodes) was applied in UK Biobank EMRs. A total of 88.6 % (3,106,440 out of 3,504,994) of records of unique ICD code per participant were successfully mapped to phecode (**Figure 15**). Among the unmatched records, 86.7% (345,426 out of 398,554 unmatched records) were ICD10 codes beginning with V (transport accidents) or Z (factors influencing health status and contact with health services), which are not expected to be directly associated with any genetic factor. In total 1,853 disease outcomes were generated with a median number of cases of 309 (range: 0 to 160,512).

Figure 15. Manipulation of electronic medical records.



I assessed the electronical medical records of participants in December 2016 and downloaded their primary diagnosis codes (the primary cause for admission in every hospitalisation episode, one for each episode), secondary diagnosis codes (other existing medical conditions, could be several for each episode) and participants' cancer registry and death registry codes through linkage with the cancer registry and death registry. Then, I treated every presence of codes as a unique line and pooled the four sources of codes we downloaded together and got a total of 7,997,465 rows of ICD codes. Subsequently, presences of codes were counted and aggregated by participants (e.g., in the previous table, 3 records of the same code for an individual were three different lines, however, after the counting, they were presented by a single line with a new column featuring the number of records of the code). I got a table of 3,504,994 rows. Finally, this table were mapped to the phecode with the phecode mapping file (https://phewascatalog.org/phecodes) (199).

5.1.4 Association between vitamin D genetic score and outcomes

I created a genetic score of 25(OH)D level by adding the number of effect alleles carried in each of the 6 SNPs and weighted based on their effect estimates from the most recent SUNLIGHT GWAS (158).

Associations between the score and potential confounding factors, including age, BMI, time spend outdoors in summer, time spend outdoors in winter, sex, UK Biobank assessment centre, average household income before tax, qualification and alcohol frequency were tested. Only UK Biobank assessment centre was statistically significantly associated with the 25(OH)D score (P = 1.30*10^{-17}) **(Table 25)**. This might be caused by the unevenly distribution of rs10741657, rs12785878 and rs10745742 between assessment centres, as had been explored in **Section 5.1.2**.

Table 25. Association of the 25(OH)D score with potential confounding factors.

	Continuous		Categorical	
Confounding factors	Beta (SE)	*P*-value	*F*-value	*P*-value
Age	0.156 (0.202)	0.441		
BMI	0.224 (0.121)	0.063		
Time spend outdoors in summer	-0.077(0.091)	0.394		
Time spend outdoors in winter	0.083(0.119)	0.485		
Sex			0.455	0.500
Assessment center			6.164	1.30*10^{-17}*
Average household income before tax			1.213	0.296
Qualification			0.490	0.843
Alcohol intake frequency			1.419	0.203

A total of 920 outcomes had a case size of greater than 200. And hence they were tested in our PheWAS analysis. No phenotypes survived Bonferroni correction (0.05/920 = 5.43*10^{-5}). There were only two phenotypes with suggestive P values smaller than 0.001, which were delirium (517 cases, P = 1.83*10^{-4}) and nephrotic syndrome (374 cases, P = 9.75*10^{-4}) (**Figure 16**). The P value for association between the score and vitamin D deficiency was 0.00116 (291 cases), which was the third

smallest *P* value among all tested associations. Although not statistically significant, it supported the validity of the instrument and of the method I applied. Finally, 52 outcomes had *P* values of lower than 0.05.

The same analysis was implemented after sex stratification (**Table 26** and **Table 27**). In females, UK Biobank assessment centre was associated significantly with the score with a *P* value of $9.06*10^{-10}$. In males, UK Biobank assessment centre was associated significantly with the score with a *P* value of $1.54*10^{-6}$. None of the other confounding factors was associated with the 25(OH)D score after sex stratification.

Table 26. Association of the 25(OH)D score with potential confounding factors in females.

	Continuous		Categorical	
Confounding factors	Beta (SE)	*P*-value	*F*-value	*P*-value
Age	0.121 (0.273)	0.658		
BMI	0.310 (0.178)	0.082		
Time spend outdoors in summer	-0.072 (0.120)	0.550		
Time spend outdoors in winter	0.045 (0.163)	0.781		
Assessment center			4.078	$9.06*10^{-10}$
Average household income before tax			1.161	0.324
Qualification			0.321	0.945
Alcohol intake frequency			0.767	0.600

Table 27. Association of the 25(OH)D score with potential confounding factors in males.

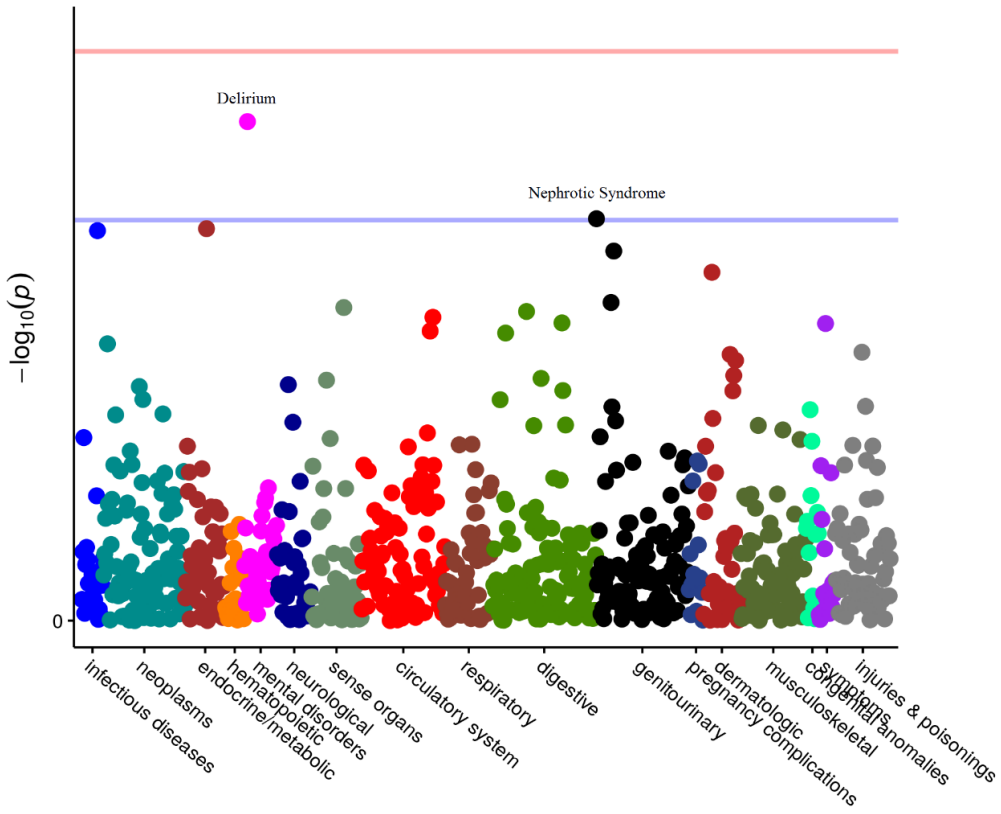
	Continuous		Categorical	
Confounding factors	Beta (SE)	*P*-value	*F*-value	*P*-value
Age	0.188 (0.300)	0.532		
BMI	0.109 (0.157)	0.487		
Time spend outdoors in summer	-0.100 (0.136)	0.460		
Time spend outdoors in winter	0.103 (0.170)	0.544		
Assessment center			3.141	$1.54*10^{-6}$
Average household income before tax			0.961	0.450
Qualification			0.668	0.700
Alcohol intake frequency			1.460	0.187

A total of 680 outcomes were tested in the PheWAS in females. Thus, after applying Bonferroni correction, a *P* value of lower than $7.35*10^{-5}$ was considered as statistically significant. None of the outcomes survived Bonferroni correction. There was only one outcome with *P* value less than 0.001, which was otitis externa (N of cases = 206; beta = -3.373; se = 0.992; $P = 6.74*10^{-4}$). A total of 26 outcomes had a *P* values less than

0.05 (**Figure 17**).

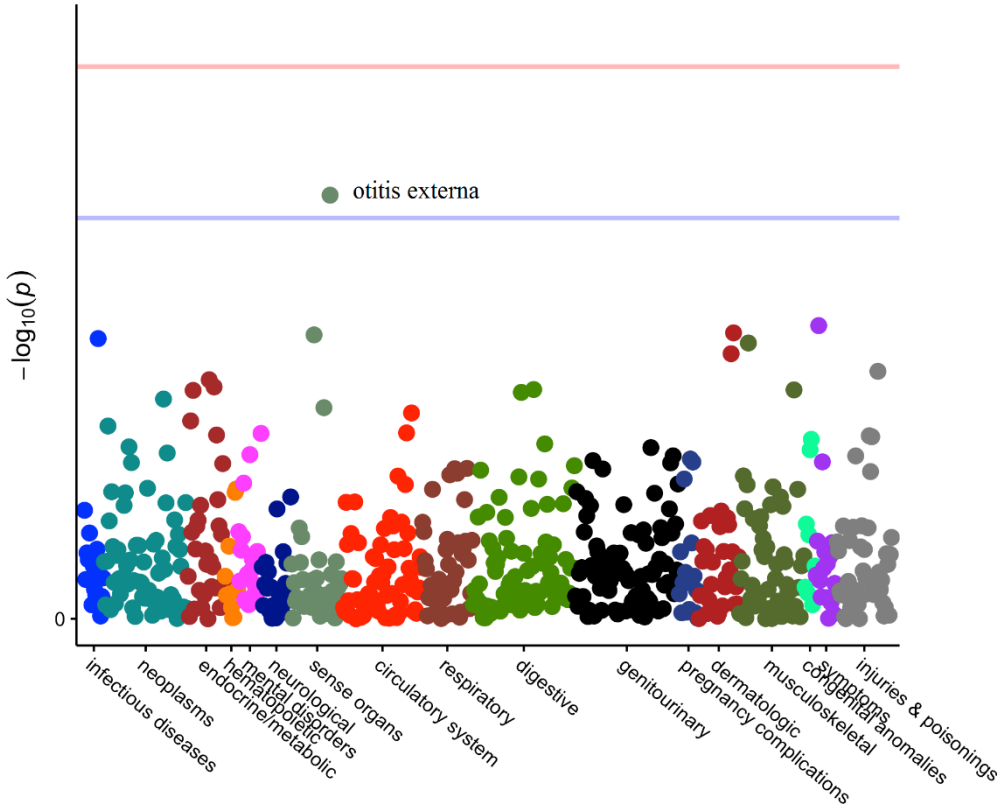
A total of 628 outcomes were tested in the PheWAS in males. Thus, applying Bonferroni correction, a *P* value of lower than $7.96*10^{-5}$ was considered as statistically significant. None of the outcomes survived Bonferroni correction. None of the tested outcomes had *P* value less than 0.001. A total of 47 outcomes had a *P* value less than 0.05 (**Figure 18**).

Figure 16. Manhattan plot for the PheWAS of 25(OH)D score.



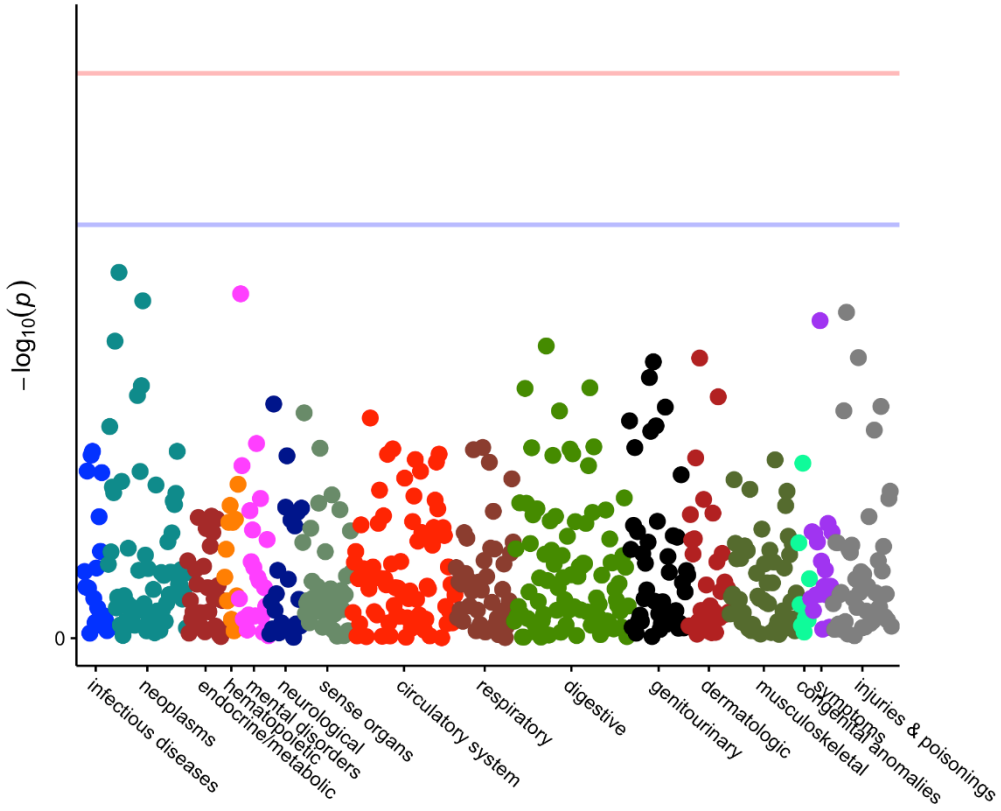
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There were only two phenotypes with *P* value less than 0.001, which were delirium ($P=1.83*10^{-4}$) and nephrotic syndrome ($P=9.75*10^{-4}$).

Figure 17. Manhattan plot for the PheWAS of 25(OH)D score in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There was only one phenotype with a *P* value less than 0.001, which was otitis externa ($P = 6.74*10^{-4}$).

Figure 18. Manhattan plot for the PheWAS of 25(OH)D score in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

5.1.5 Association between single genetic variants and outcomes

Following the PheWAS using score of all 6 SNPs, I also conducted PheWAS for every SNP among the 6 25(OH)D level related SNPs.

PheWAS for rs3755967 (GC)

The genotype of rs3755967 was not associated with any potential confounding factors (**Table 28**). It was not associated with confounding factors in neither females nor males after sex stratification (**Table 29** and **Table 30**).

In PheWAS in both genders, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest *P* value was pilonidal cyst (beta = -0.290, se = 0.077, $P = 1.53*10^{-4}$), which was the only outcome with *P* value less than 0.001. Vitamin D deficiency was the outcome with the second lowest *P* value, which was close to 0.001 (beta = 0.286, se = 0.087, *P* = 0.001). A total of 49 outcomes were with *P* values less than 0.05 (**Figure 19**). In the PheWAS in females, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest *P* value was otitis externa (beta = 0.317, se = 0.103, *P* = 0.002). There were 28 outcomes with *P* values less than 0.05 (**Figure 20**). In the PheWAS in males, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest *P* value was pilonidal cyst (beta = -0.298, se = 0.090, *P* = 0.001), which was the only outcome with *P* value less than 0.001. In total, there were 34 outcomes with *P* value less than 0.05 (**Figure 21**).

Table 28. Association of rs3755967 genotype with potential confounding factors.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	1.024	0.359			
BMI	1.641	0.194			
Time spend outdoors in summer	0.447	0.640			
Time spend outdoors in winter	0.756	0.469			
Sex			5.140	2	0.077

Assessment center	53.766	42	0.105
Average household income before tax	7.478	12	0.825
Qualification	17.384	14	0.236
Alcohol intake frequency	15.137	12	0.234

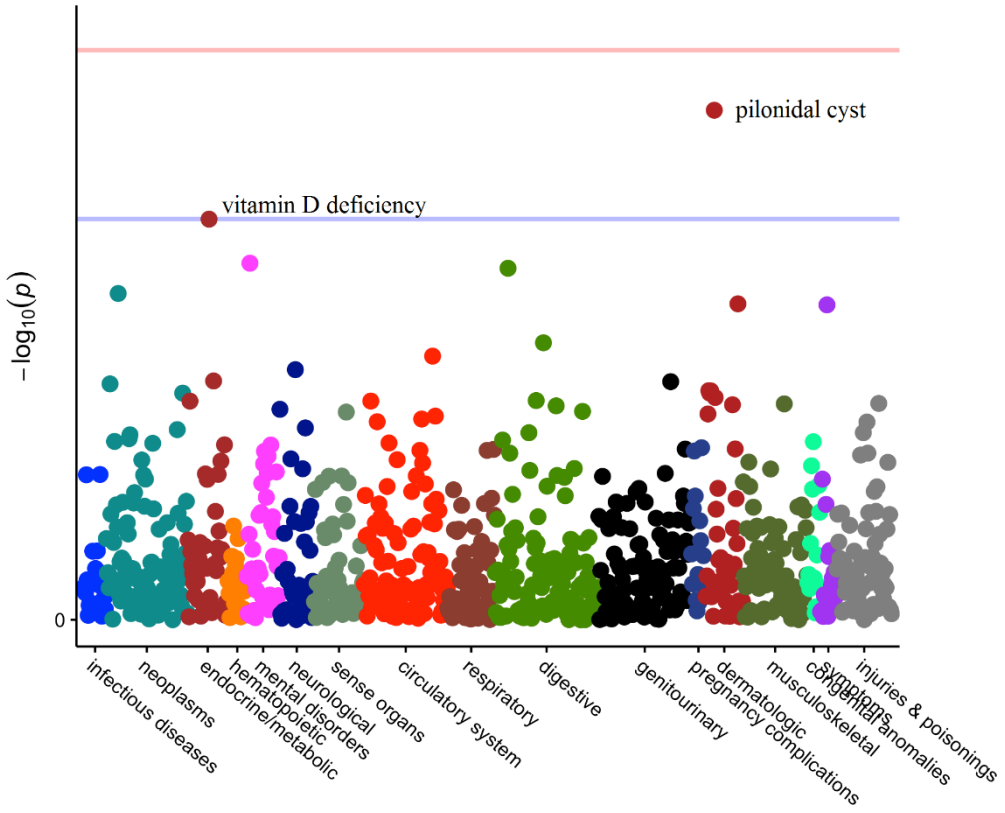
Table 29. Association of rs3755967 genotype with potential confounding factors in females.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.656	0.519			
BMI	0.522	0.594			
Time spend outdoors in summer	0.272	0.762			
Time spend outdoors in winter	0.934	0.393			
Assessment center			43.055	42	0.426
Average household income before tax			10.066	12	0.610
Qualification			12.285	14	0.583
Alcohol intake frequency			14.135	12	0.292

Table 30. Association of rs3755967 genotype with potential confounding factors in males.

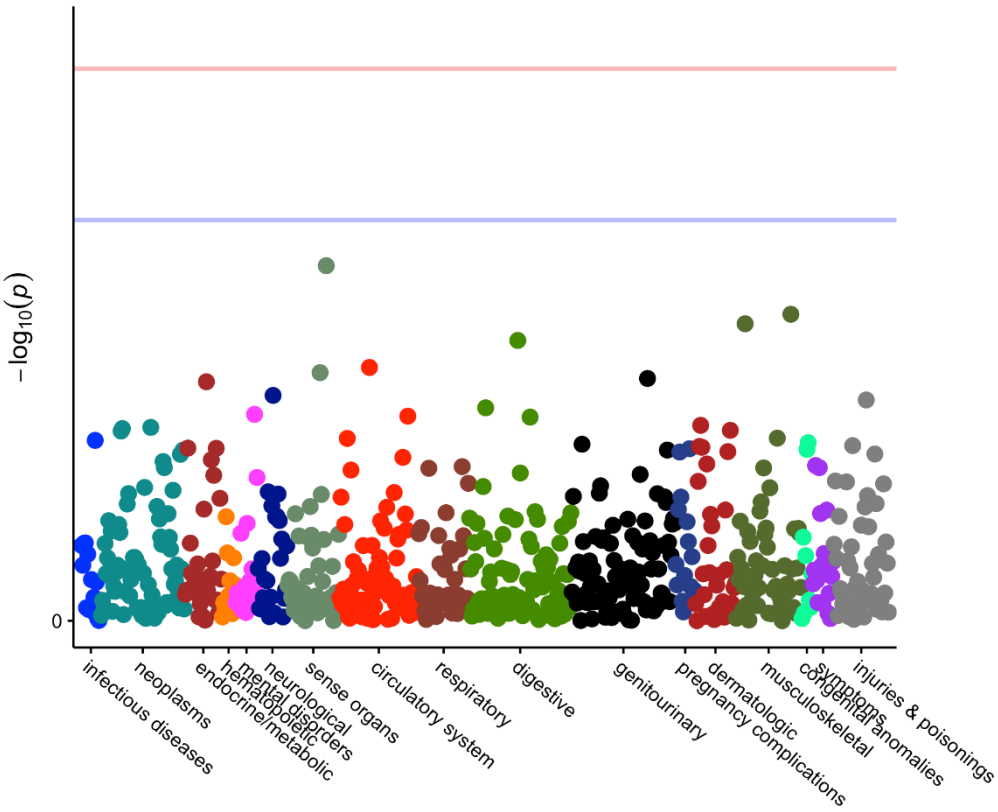
	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.350	0.705			
BMI	2.874	0.056			
Time spend outdoors in summer	0.376	0.686			
Time spend outdoors in winter	0.335	0.716			
Assessment center			48.585	42	0.225
Average household income before tax			5.132	12	0.953
Qualification			18.541	14	0.183
Alcohol intake frequency			13.258	12	0.351

Figure 19. Manhattan plot for the PheWAS of rs3755967.



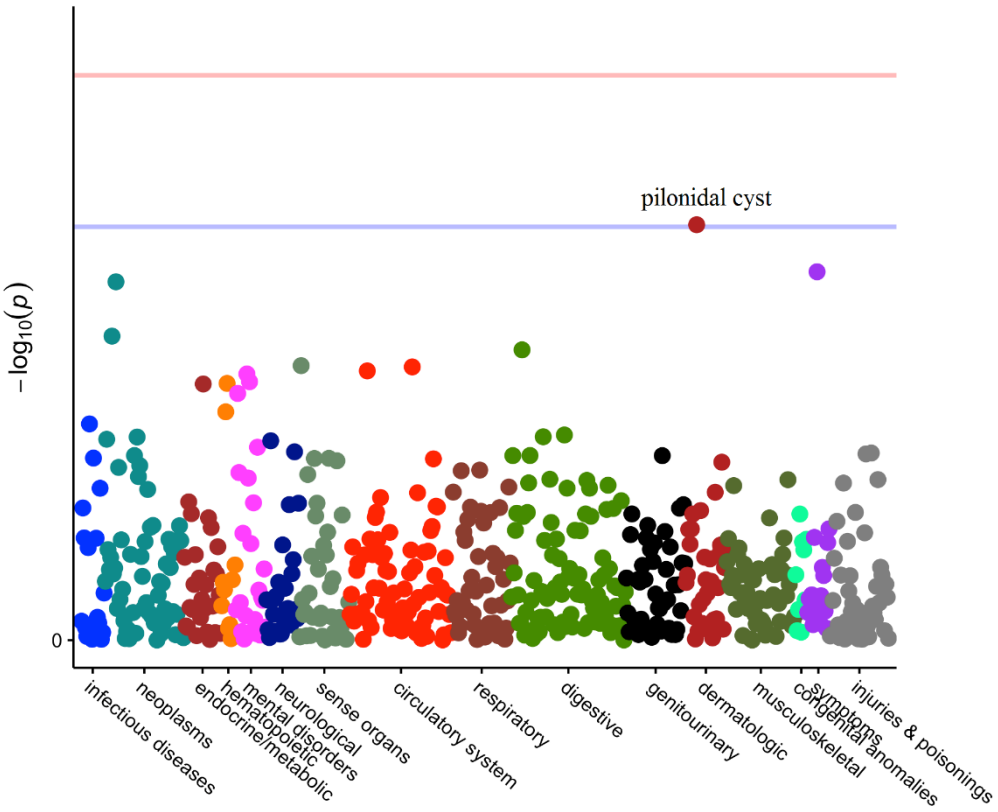
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There was only one phenotype with a *P* value less than 0.001, which was pilonidal cyst ($P=1.53*10^{-4}$). Phenotype vitamin D deficiency was with the second lowest *P* value, which was close to 0.001 ($P = 0.001$).

Figure 20. Manhattan plot for the PheWAS of rs3755967 in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 21. Manhattan plot for the PheWAS of rs3755967 in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There was only one phenotypes with a *P* value less than 0.001, which was pilonidal cyst (*P* = 0.001).

PheWAS for rs10741657 (CYP2R1)

The genotype of rs10741657 was associated with BMI (P = 0.038), gender (P = 0.036) and assessment centre (P = 0.003) (**Table 31**). It was not associated with any confounding factors, including BMI (P = 0.212 in females; P = 0.156 in males) and assessment centre (P = 0.234 in females; P = 0.106 in males) after sex stratification (**Table 32** and **Table 33**).

In PheWAS in both genders, none of the tested outcomes survived Bonferroni correction. Genotypes of rs10741657 was associated with nephrotic syndrome (beta = -0.277, se = .073, $P = 1.53*10^{-4}$), labyrinthitis (beta = -0.204, se = 0.057, $P = 3.82*10^{-4}$), and complications of cardiac/vascular device, implant, and graft (beta = -0.123, se = 0.037, P = $9.16*10^{-4}$) at P level of lower than 0.001. A total of 52 outcomes were with P values less than 0.05 (**Figure 22**). In the PheWAS in females, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest P value was diverticulosis (beta = 0.043, se = 0.014, P = 0.002). There were 42 outcomes with P values less than 0.05 (**Figure 23**). In the PheWAS in males, none of the tested outcomes survived Bonferroni correction. Premature beats (beta = -0.324, se = 0.094, $P = 5.84*10^{-4}$) and carcinoma in situ of skin (beta = -0.192, se = 0.057, $P = 8.34*10^{-4}$) were with P value less than 0.001. In total, there were 41 outcomes with P values of lower than 0.05 (**Figure 24**).

Table 31. Association of rs10741657 genotype with potential confounding factors.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.276	0.759			
BMI	3.279	0.038			
Time spend outdoors in summer	0.914	0.401			
Time spend outdoors in winter	0.077	0.926			
Sex			6.624	2	0.036
Assessment center			71.099	42	0.003

Average household income before tax			12.697	12	0.391
Qualification			15.003	14	0.378
Alcohol intake frequency			6.838	12	0.868

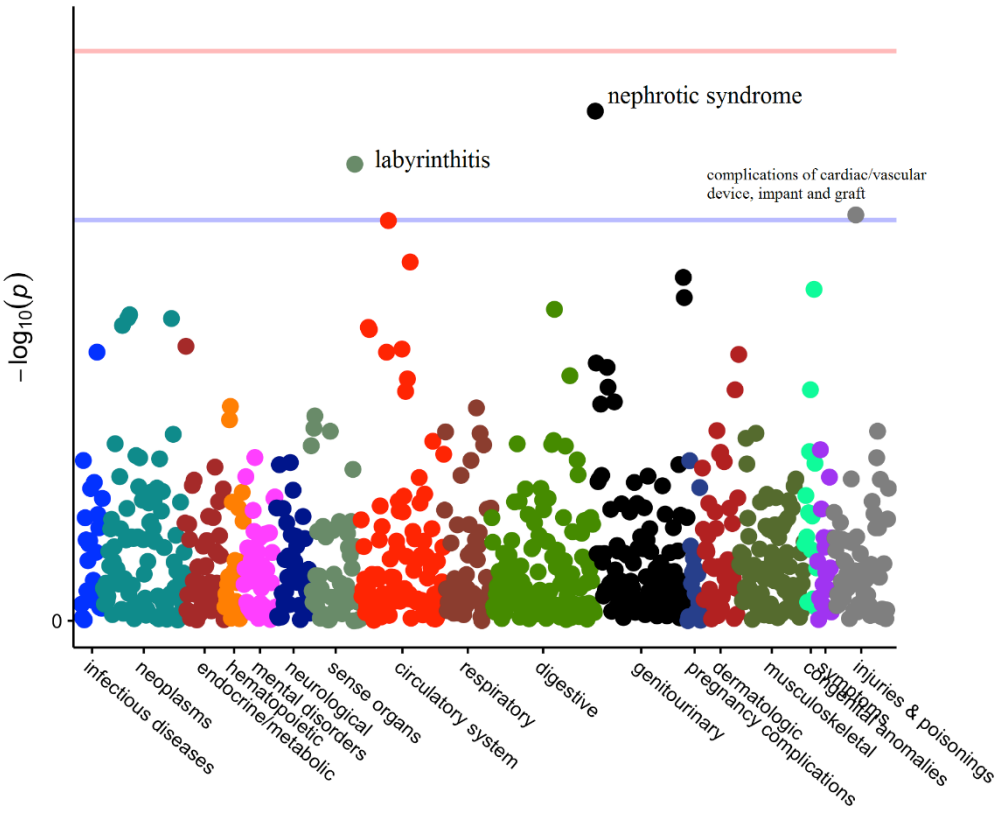
Table 32. Association of rs10741657 genotype with potential confounding factors in females.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.164	0.849			
BMI	1.549	0.212			
Time spend outdoors in summer	0.126	0.882			
Time spend outdoors in winter	0.889	0.411			
Assessment center			48.28	42	0.234
Average household income before tax			9.199	12	0.686
Qualification			13.02	14	0.525
Alcohol intake frequency			8.072	12	0.779

Table 33. Association of rs10741657 genotype with potential confounding factors in males.

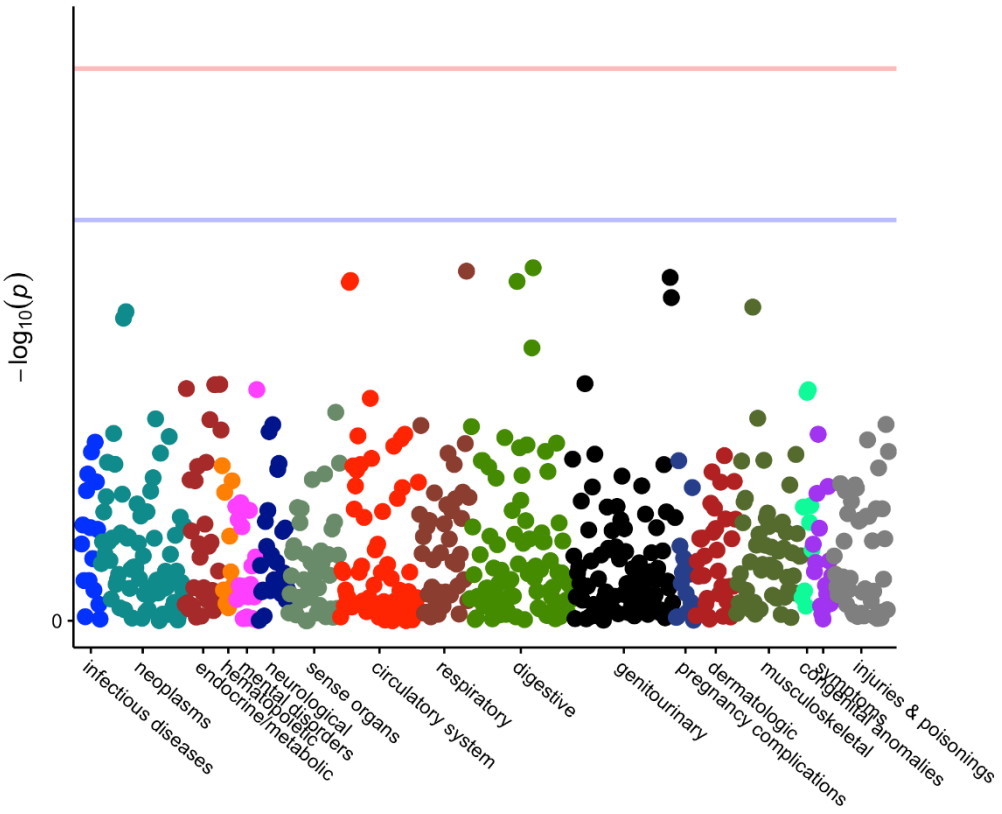
	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.210	0.811			
BMI	1.857	0.156			
Time spend outdoors in summer	2.188	0.112			
Time spend outdoors in winter	0.653	0.521			
Assessment center			53.757	42	0.106
Average household income before tax			15.616	12	0.210
Qualification			11.145	14	0.675
Alcohol intake frequency			10.691	12	0.556

Figure 22. Manhattan plot for the PheWAS of rs10741657.



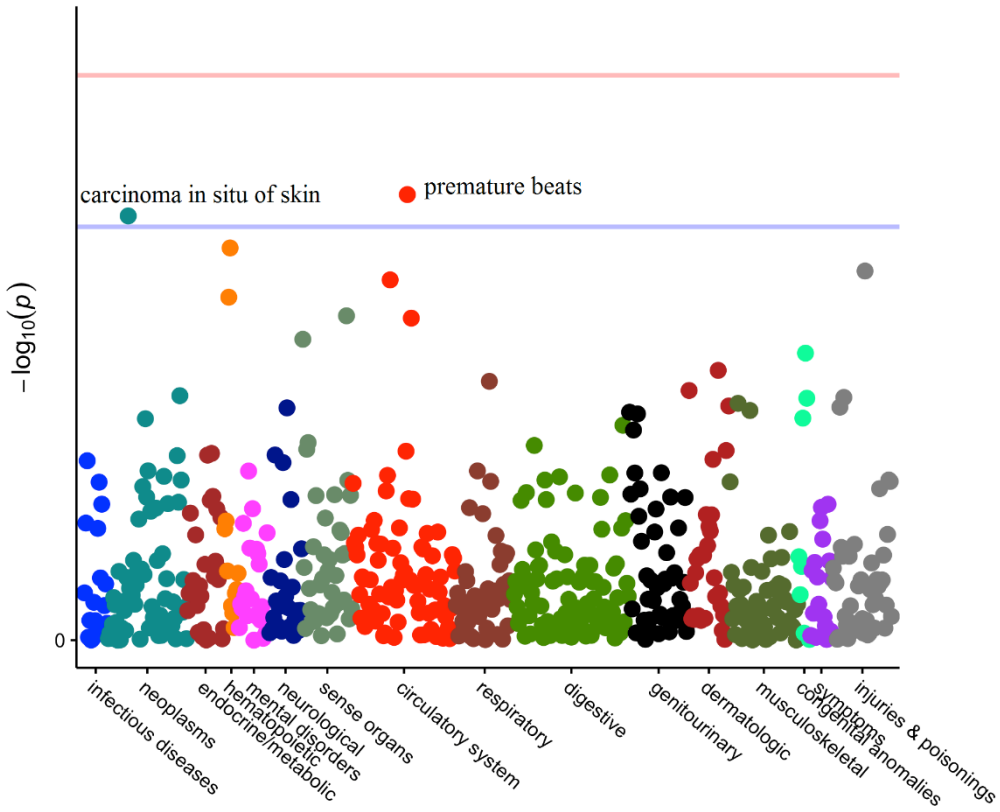
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There were three phenotypes with *P* value less than 0.001, which were nephrotic syndrome ($P = 1.53*10^{-4}$), labyrinthitis ($P = 3.82*10^{-4}$) and complications of cardiac/vascular device, implant and graft ($P = 9.16*10^{-4}$).

Figure 23. Manhattan plot for the PheWAS of rs10741657 in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 24. Manhattan plot for the PheWAS of rs10741657 in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There were two phenotypes with *P* value less than 0.001, which were premature beats ($P = 5.84*10^{-4}$) and carcinoma in situ of skin ($P = 8.34*10^{-4}$).

PheWAS for rs12785878 (DHCR7)

The genotypes of rs12785878 was association with educational qualifications (P = 0.046) and UK Biobank assessment centre ($P < 2.2*10^{-16}$) (**Table 34**). In females, rs12785878 was associated with qualification (P = 0.024) and UK Biobank assessment centre ($P < 2.2*10^{-16}$) (**Table 35**). In males, rs12785878 was associated with UK Biobank assessment centre ($P < 2.2*10^{-16}$), but not qualifications (P = 0.621) (**Table 36**).

In PheWAS in both genders, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest P value was visual disturbances (beta = 0.174, se = 0.057, P = 0.002). There were 46 outcomes with P value less than 0.05 altogether (**Figure 25**). In the PheWAS in females, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest P value was sciatica (beta = -0.140, se = 0.053, P = 0.009). There were 35 outcomes with P value less than 0.05 in total (**Figure 26**). In the PheWAS in males, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest P value was internal derangement of the knee (beta = 0.064, se = 0.021, P = 0.002). There were 42 outcomes with P value less than 0.05 in total (**Figure 27**).

Table 34. Association of rs12785878 genotype with potential confounding factors.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	1.662	0.190			
BMI	2.822	0.059			
Time spend outdoors in summer	0.625	0.535			
Time spend outdoors in winter	0.040	0.960			
Sex			0.498	2	0.780
Assessment center			338.2	42	$< 2.2*10^{-16}$
Average household income before tax			18.47	12	0.102
Qualification			24.023	14	0.046
Alcohol intake frequency			14.366	12	0.278

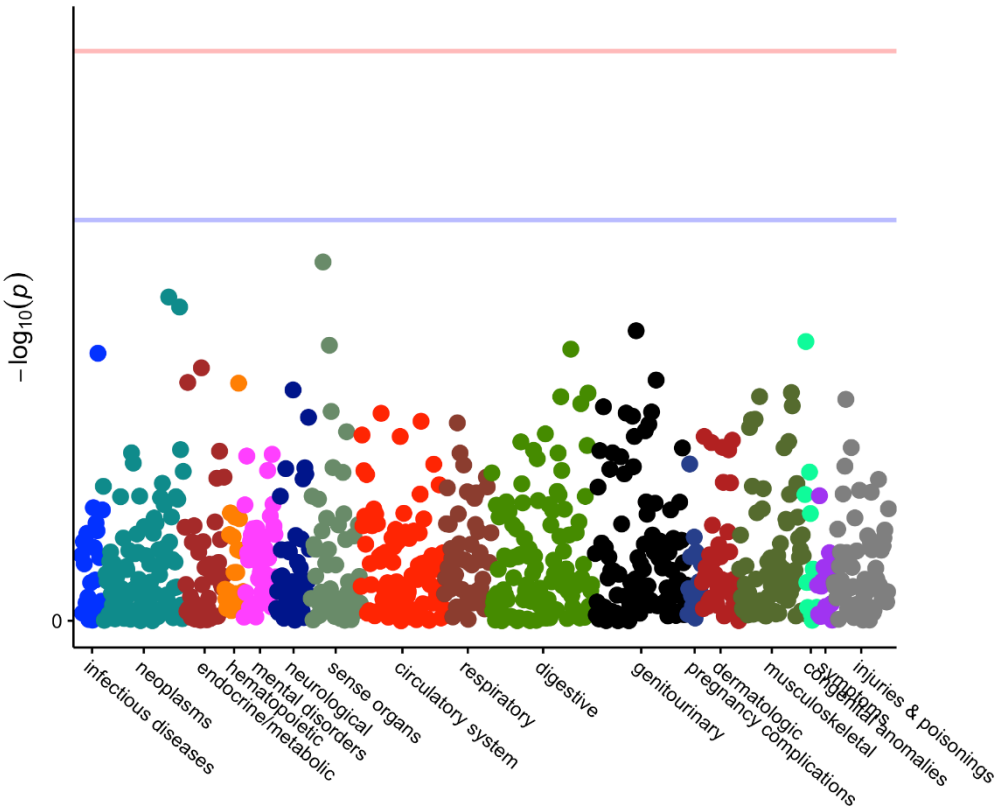
Table 35. Association of rs12785878 genotype with potential confounding factors in females.

Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	0.395	0.673			
BMI	2.684	0.068			
Time spend outdoors in summer	0.333	0.717			
Time spend outdoors in winter	0.527	0.590			
Assessment center			192.67	42	$< 2.2*10^{-16}$
Average household income before tax			21.038	12	0.050
Qualification			26.199	14	0.024
Alcohol intake frequency			14.812	12	0.252

Table 36. Association of rs12785878 genotype with potential confounding factors in males.

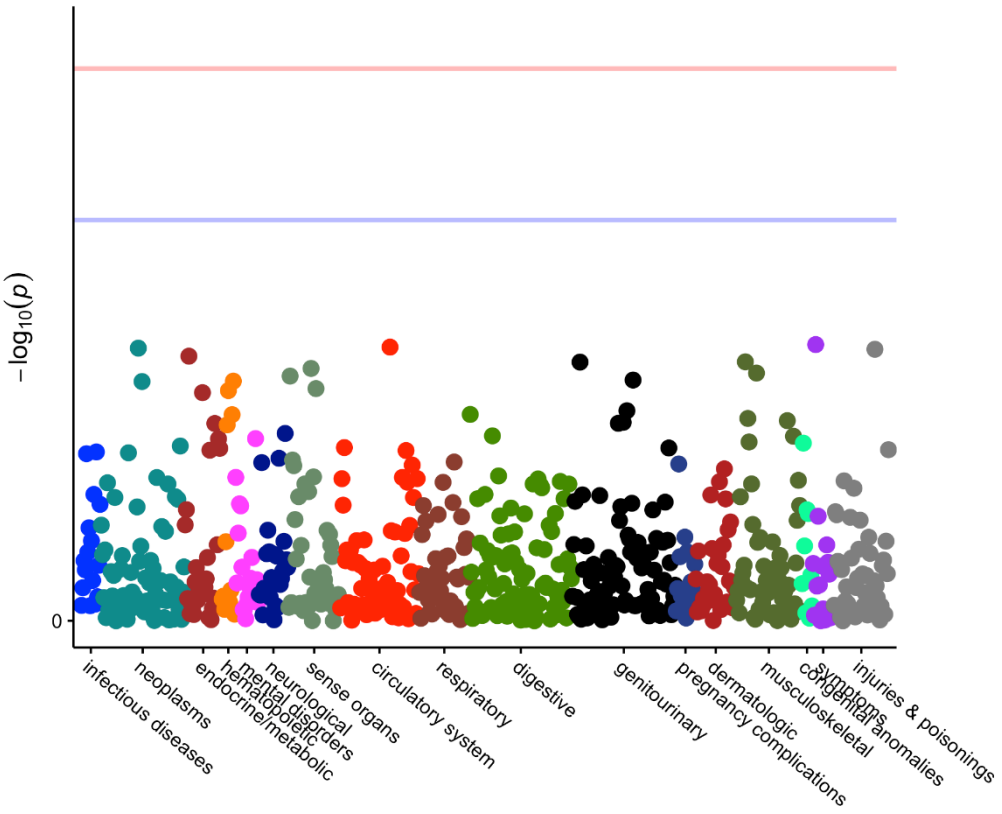
Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	1.514	0.220			
BMI	0.527	0.590			
Time spend outdoors in summer	0.602	0.548			
Time spend outdoors in winter	0.495	0.610			
Assessment center			184.51	42	$< 2.2*10^{-16}$
Average household income before tax			12.042	12	0.442
Qualification			11.814	14	0.621
Alcohol intake frequency			20.651	12	0.056

Figure 25. Manhattan plot for the PheWAS of rs12785878.



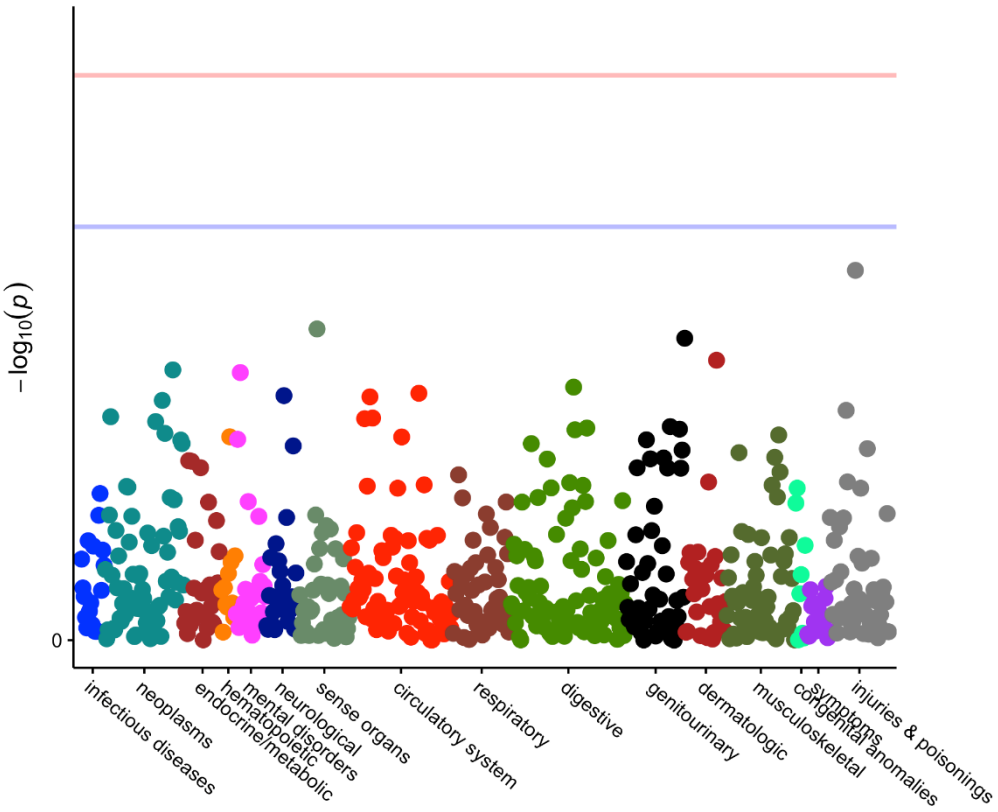
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 26. Manhattan plot for the PheWAS of rs12785878 in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 27. Manhattan plot for the PheWAS of rs12785878 in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

PheWAS for rs10745742 (AMDHD1)

The genotype of rs10745742 was associated with UK Biobank assessment centre ($P < 2.2*10^{-16}$) (**Table 37**). It was associated with assessment centre in both males and female ($P = 2.25*10^{-13}$ in females; $P = 1.18*10^{-9}$ in males) after sex stratification (**Table 38** and **Table 39**). The genotype of rs10745742 was not associated with any other confounding factors.

In PheWAS for both genders, none of the tested outcomes survived Bonferroni correction. "Allied disorders of spine" was the outcome with the lowest *P* value (beta = 0.050, se = 0.016, *P* = 0.001). In total, there were 56 outcomes with *P* value less than 0.05 (**Figure 28**). In the PheWAS in females, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest *P* value was congestive heart failure (beta = 0.200, se = 0.067, *P* = 0.003). There were 34 outcomes with *P* value less than 0.05 (**Figure 29**). In the PheWAS in males, none of the tested outcomes survived Bonferroni correction. "Allied disorders of spine" was the outcome with the lowest *P* value (beta = 0.076, se = 0.024, *P* = 0.001). There were 28 outcomes with *P* value less than 0.05 (**Figure 30**).

Table 37. Association of rs10745742 genotype with potential confounding factors.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.260	0.771			
BMI	0.497	0.609			
Time spend outdoors in summer	0.301	0.740			
Time spend outdoors in winter	0.069	0.933			
Sex			4.757	2	0.093
Assessment center			223.39	42	$< 2.2*10^{-16}$
Average household income before tax			10.321	12	0.588
Qualification			5.947	14	0.968
Alcohol intake frequency			11.022	12	0.527

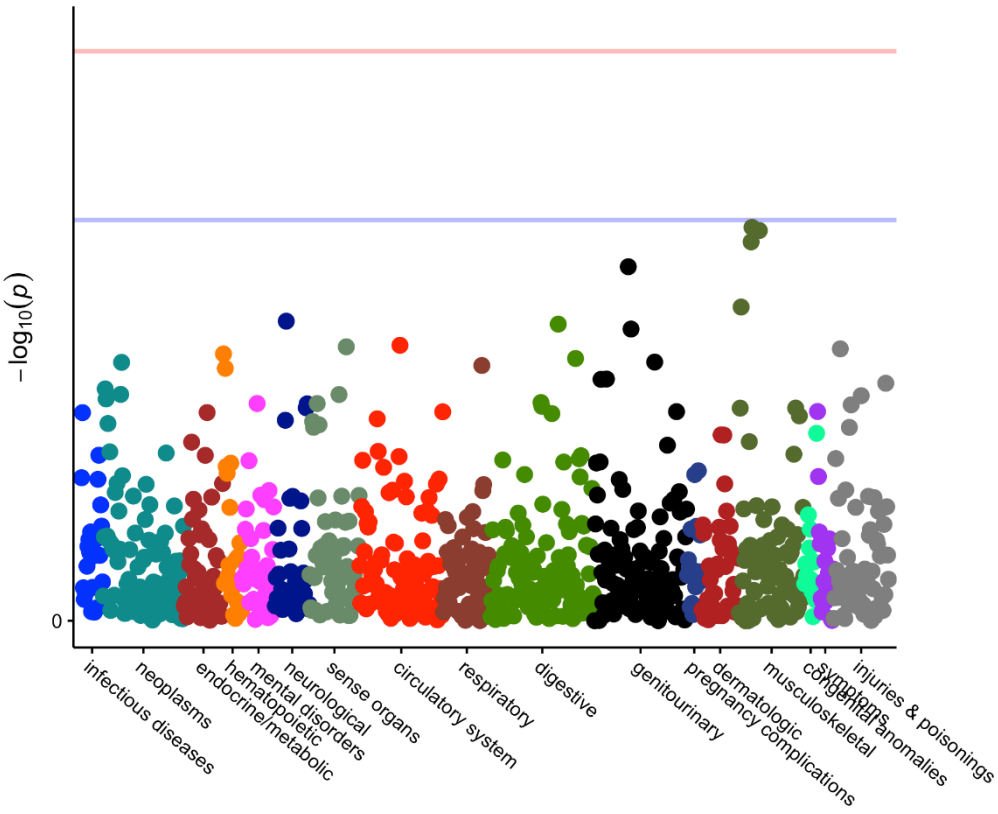
Table 38. Association of rs10745742 genotype with potential confounding factors in females.

Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	0.797	0.451			
BMI	0.887	0.412			
Time spend outdoors in summer	1.296	0.274			
Time spend outdoors in winter	0.690	0.502			
Assessment center			145.75	42	$2.25*10^{-13}$
Average household income before tax			8.847	12	0.716
Qualification			6.617	14	0.949
Alcohol intake frequency			8.157	12	0.773

Table 39. Association of rs10745742 genotype with potential confounding factors in males.

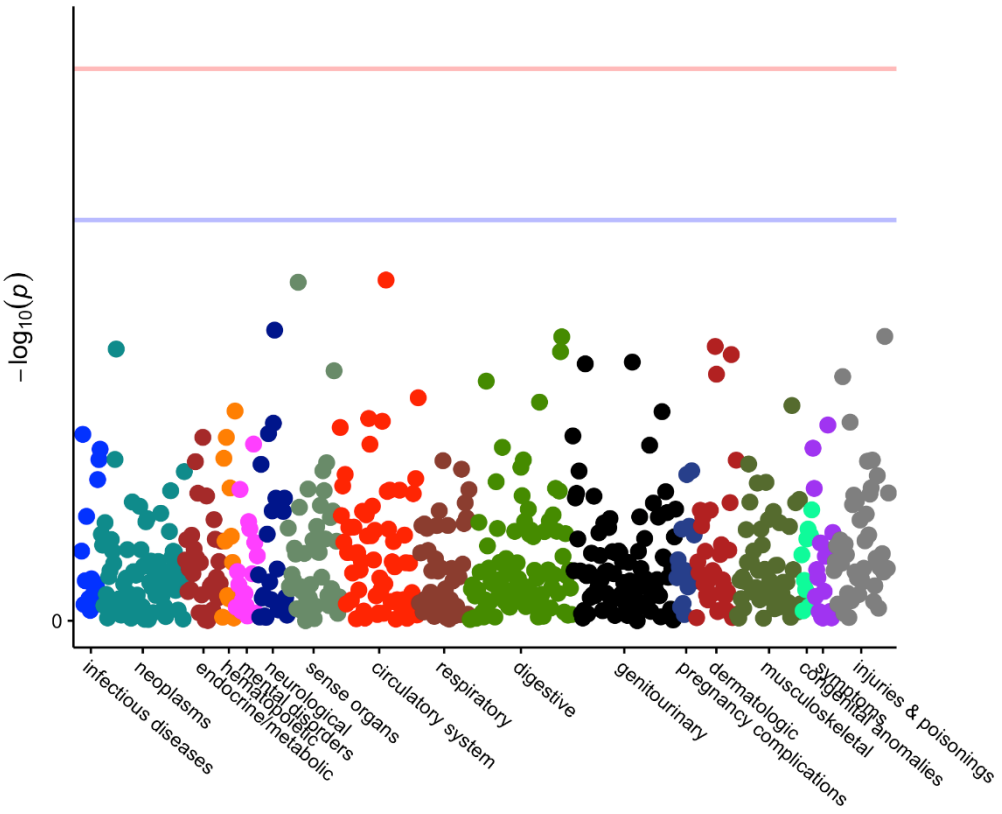
Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	0.913	0.401			
BMI	0.350	0.705			
Time spend outdoors in summer	0.599	0.550			
Time spend outdoors in winter	0.909	0.403			
Assessment center			121.5	42	$1.18*10^{-9}$
Average household income before tax			9.022	12	0.701
Qualification			7.367	14	0.920
Alcohol intake frequency			11.084	12	0.522

Figure 28. Manhattan plot for the PheWAS of rs10745742.



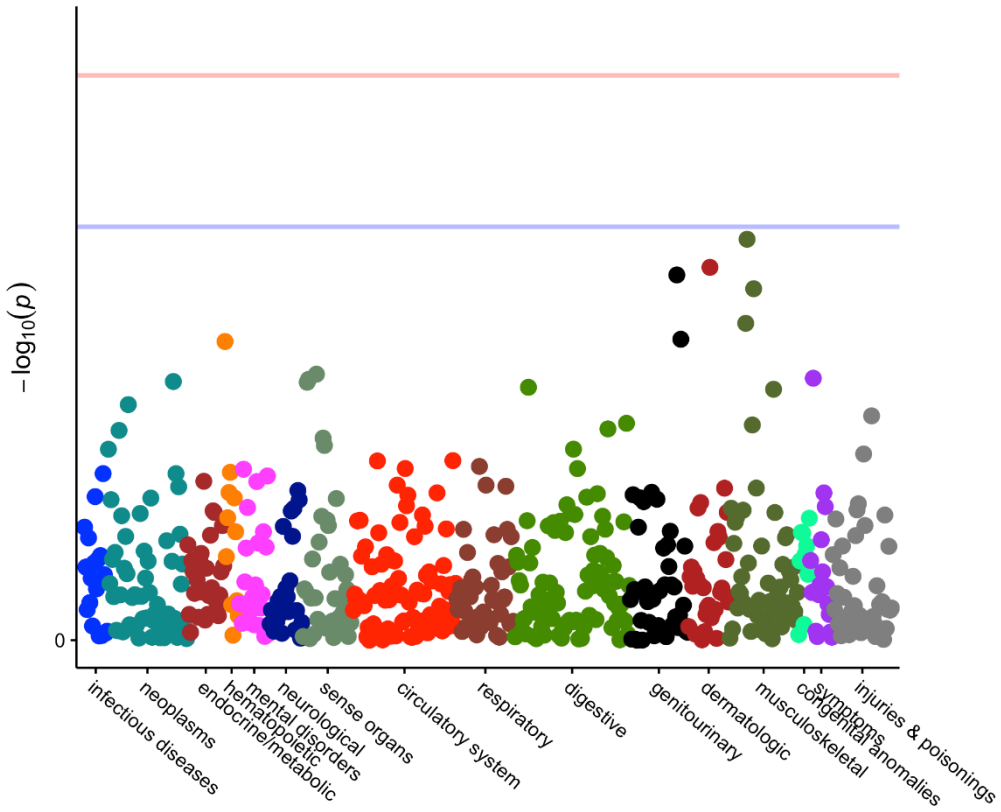
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 29. Manhattan plot for the PheWAS of rs10745742 in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 30. Manhattan plot for the PheWAS of rs10745742 in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

PheWAS for rs8018720 (SEC23A)

The genotype of rs8018720 was not associated with any confounding factors in general (**Table 40**). It was not associated with any confounding factors in either females nor males after sex stratification (**Table 41** and **Table 42**).

In PheWAS in both genders, none of the outcomes survived Bonferroni correction. The genotype was associated with "myalgia and myositis" (beta = 0.331, se = 0.094, $P = 4.02*10^{-4}$) and "rheumatic disease of the heart valves" (beta = -0.225, se = 0.064, $P = 4.21*10^{-4}$) at *P* value less than 0.001. In total, there were 50 outcomes with a *P* value less than 0.05 (**Figure 31**). In the PheWAS in females, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest *P* value was "rheumatic disease of the heart valves" (beta = -0.318, se = 0.101, *P* = 0.002). There were 29 outcomes with *P* value less than 0.05 (**Figure 32**). In the PheWAS in males, none of the tested outcomes survived Bonferroni correction. The outcome with the lowest *P* value was "staphylococcus infections" (beta = -0.151, se = 0.047, *P* = 0.001). All outcomes were associated with rs8018720 at *P* levels of greater than 0.001. There were 30 outcomes with *P* value less than 0.05 (**Figure 33**).

Table 40. Association of rs8018720 genotype with potential confounding factors.

Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	0.567	0.567			
BMI	0.732	0.481			
Time spend outdoors in summer	1.379	0.252			
Time spend outdoors in winter	1.556	0.211			
Sex			5.476	2	0.065
Assessment center			49.326	42	0.204
Average household income before tax			13.237	12	0.352
Qualification			10.316	14	0.739
Alcohol intake frequency			10.097	12	0.608

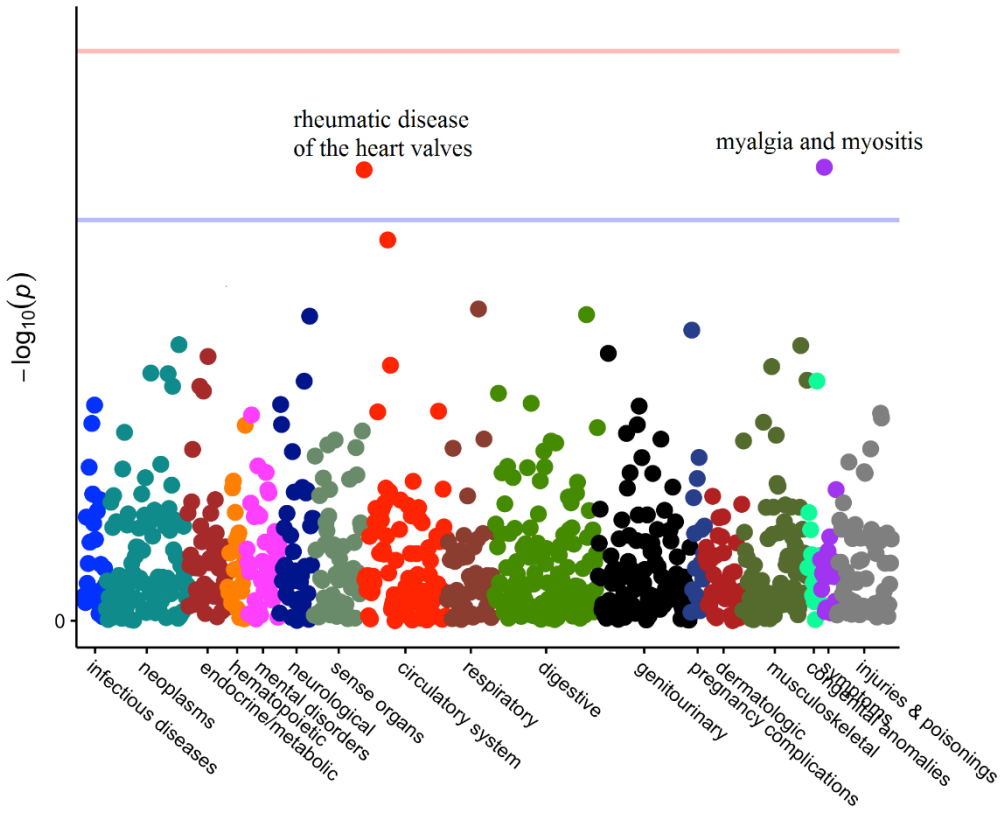
Table 41. Association of rs8018720 genotype with potential confounding factors in females.

Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	0.569	0.566			
BMI	0.173	0.841			
Time spend outdoors in summer	1.677	0.187			
Time spend outdoors in winter	2.146	0.117			
Assessment center			45.241	42	0.338
Average household income before tax			6.746	12	0.874
Qualification			10.207	14	0.747
Alcohol intake frequency			6.138	12	0.909

Table 42. Association of rs8018720 genotype with potential confounding factors in males.

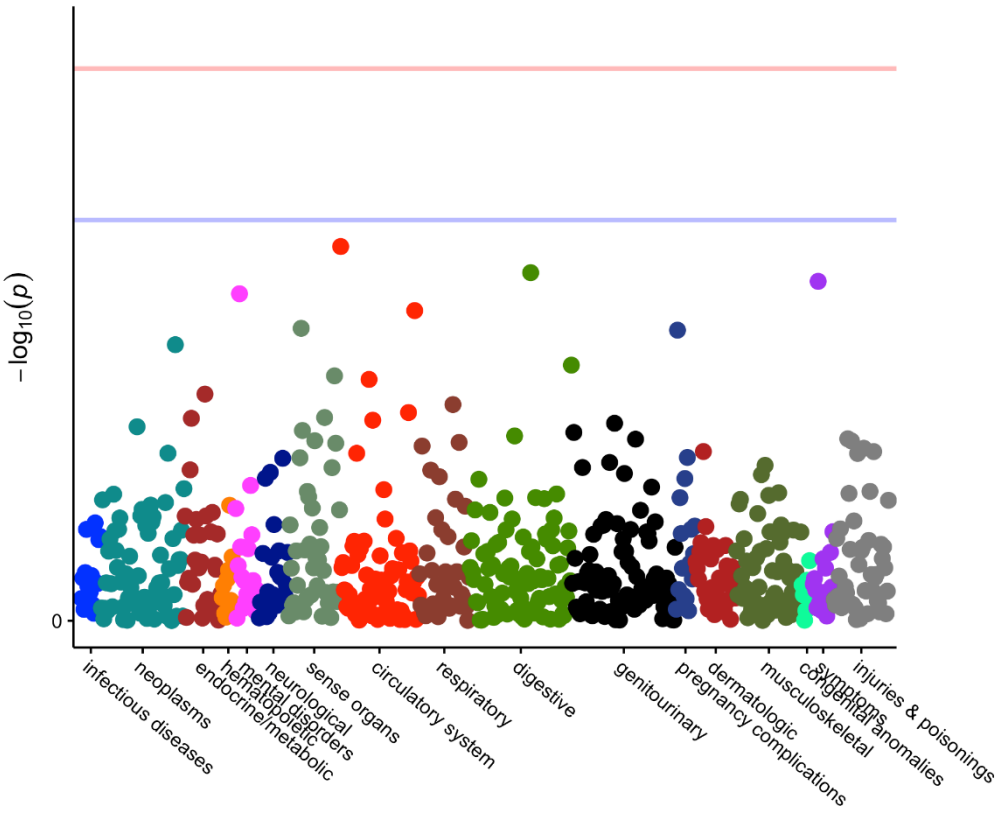
Confounding factors	Continuous		Categorical		
	F-value	*P*-value	X^2	*df*	*P*-value
Age	0.128	0.880			
BMI	1.323	0.266			
Time spend outdoors in summer	0.424	0.654			
Time spend outdoors in winter	0.415	0.660			
Assessment center			44.61	42	0.363
Average household income before tax			13.448	12	0.337
Qualification			10.124	14	0.753
Alcohol intake frequency			16.465	12	0.171

Figure 31. Manhattan plot for the PheWAS of rs8018720.



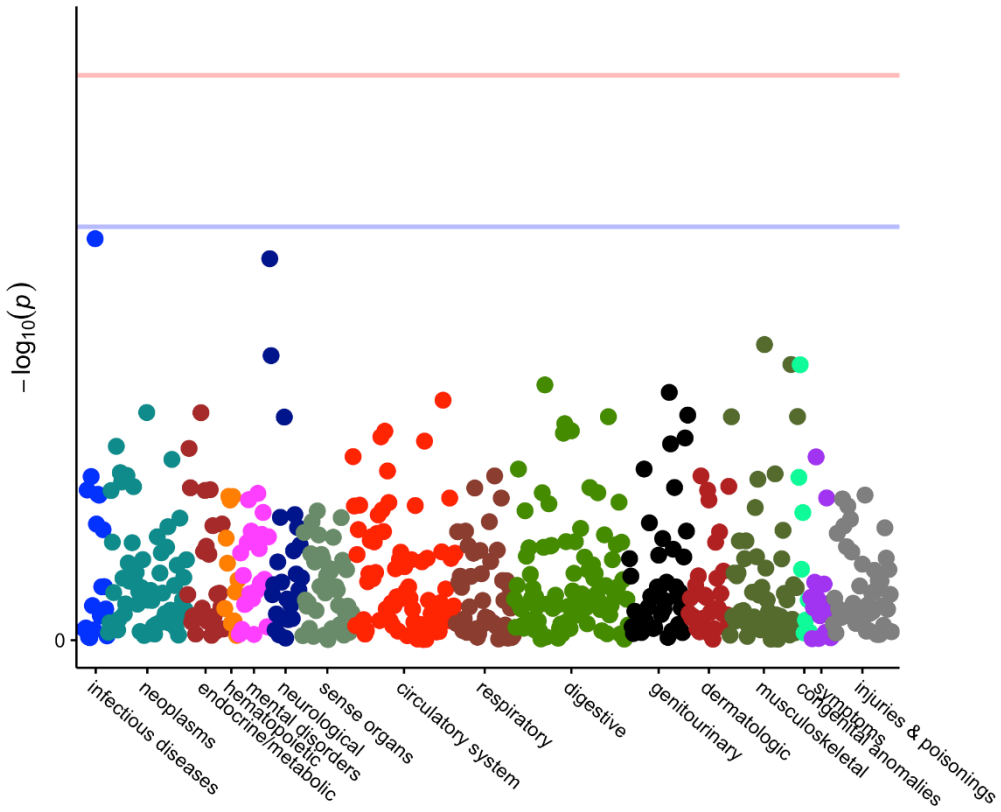
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There were two phenotypes with *P* value less than 0.001, which were myalgia and myositis ($P = 4.02*10^{-4}$) and rheumatic disease of the heart valves ($P = 4.21*10^{-4}$).

Figure 32. Manhattan plot for the PheWAS of rs8018720 in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

Figure 33. Manhattan plot for the PheWAS of rs8018720 in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction or with *P* value less than 0.001.

PheWAS for rs17216707 (CYP24A1)

The genotype of rs17216707 was not associated with any confounding factors in general (**Table 43**). It was associated with alcohol intake frequency in males after sex stratification ($P = 0.022$) (**Table 45**). It was not associated with any others confounding factors in females or males after sex stratification (**Table 44** and **Table 45**).

In PheWAS in both genders, associations between the genotype of rs17216707 and "calculus of ureter" (beta = -0.219, se = 0.045, $P = 1.14*10^{-6}$), "urinary calculus" (beta = -0.129, se = 0.027, $P = 1.31*10^{-6}$), "alveolar and parieto-alveolar pneumonopathy" (beta = 0.418, se = 0.101, $P = 3.53*10^{-5}$) survived Bonferroni correction. In addition, "calculus of kidney" (beta = -0.139, se = 0.038, $P = 2.98*10^{-4}$) was associated with rs17216707 at *P* level less than 0.001. There were 62 outcomes with *P* value less than 0.05 (**Figure 34**). In PheWAS in females, none of the tested outcomes survived Bonferroni correction. The genotype of rs17216707 was associated with "calculus of kidney" (beta = -0.270, se = 0.071, $P = 1.39*10^{-4}$), "urinary calculus" (beta = -0.175, se = 0.050, $P = 4.35*10^{-4}$) and "calculus of ureter" (beta = -0.333, se = 0.096, $P = 5.51*10^{-4}$) at a *P* level less than 0.001. There were 37 outcomes with a *P* value less than 0.05 (**Figure 35**). In PheWAS in males, none of the tested outcomes survived Bonferroni correction. Genotype of rs17216707 was associated with "calculus of ureter" (beta = -0.185, se = 0.051, P = $2.75*10^{-4}$) and "urinary calculus" (beta = -0.110, se = 0.032, P = $5.08*10^{-4}$) at *P* level less than 0.001. There were 48 outcomes with a *P* value less than 0.05 (**Figure 36**).

Table 43. Association of rs17216707 genotype with potential confounding factors.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.827	0.437			
BMI	1.095	0.335			
Time spend outdoors in summer	0.180	0.835			
Time spend outdoors in winter	0.458	0.632			
Sex			1.778	2	0.411

Assessment center	29.873	42	0.920
Average household income before tax	11.128	12	0.518
Qualification	16.363	14	0.292
Alcohol intake frequency	18.887	12	0.091

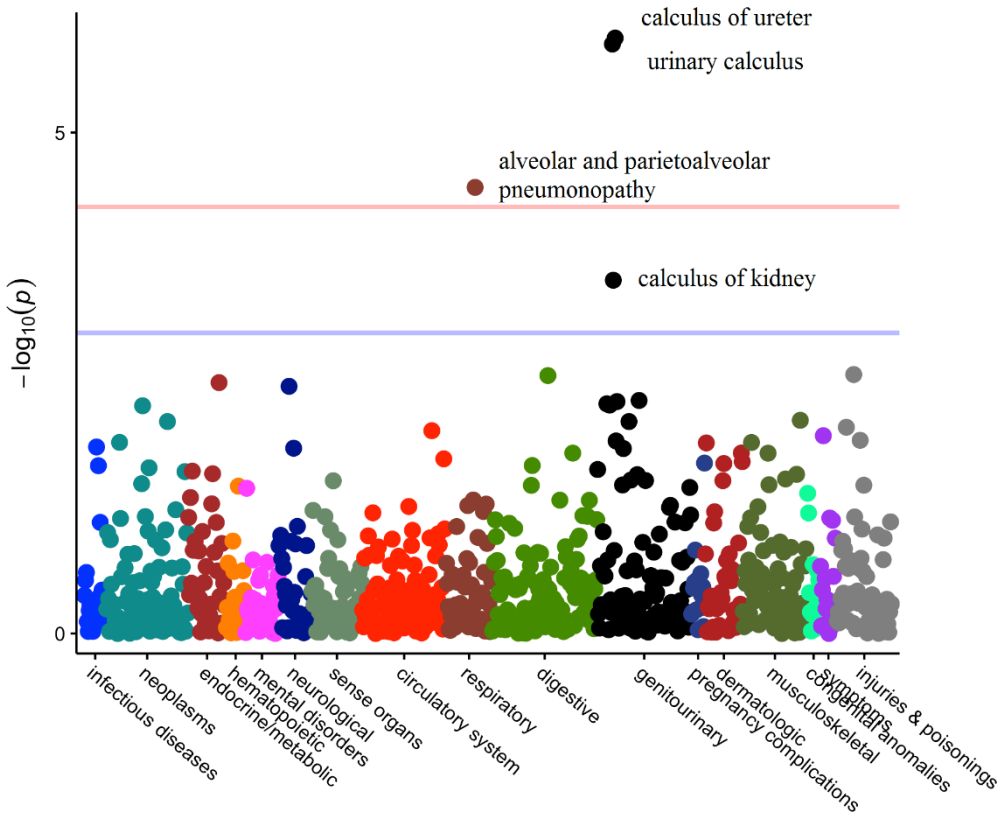
Table 44. Association of rs17216707 genotype with potential confounding factors in females.

	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	1.971	0.139			
BMI	0.619	0.539			
Time spend outdoors in summer	0.329	0.720			
Time spend outdoors in winter	0.281	0.755			
Assessment center			35.515	42	0.750
Average household income before tax			14.897	12	0.247
Qualification			9.656	14	0.787
Alcohol intake frequency			9.60	12	0.651

Table 45. Association of rs17216707 genotype with potential confounding factors in males.

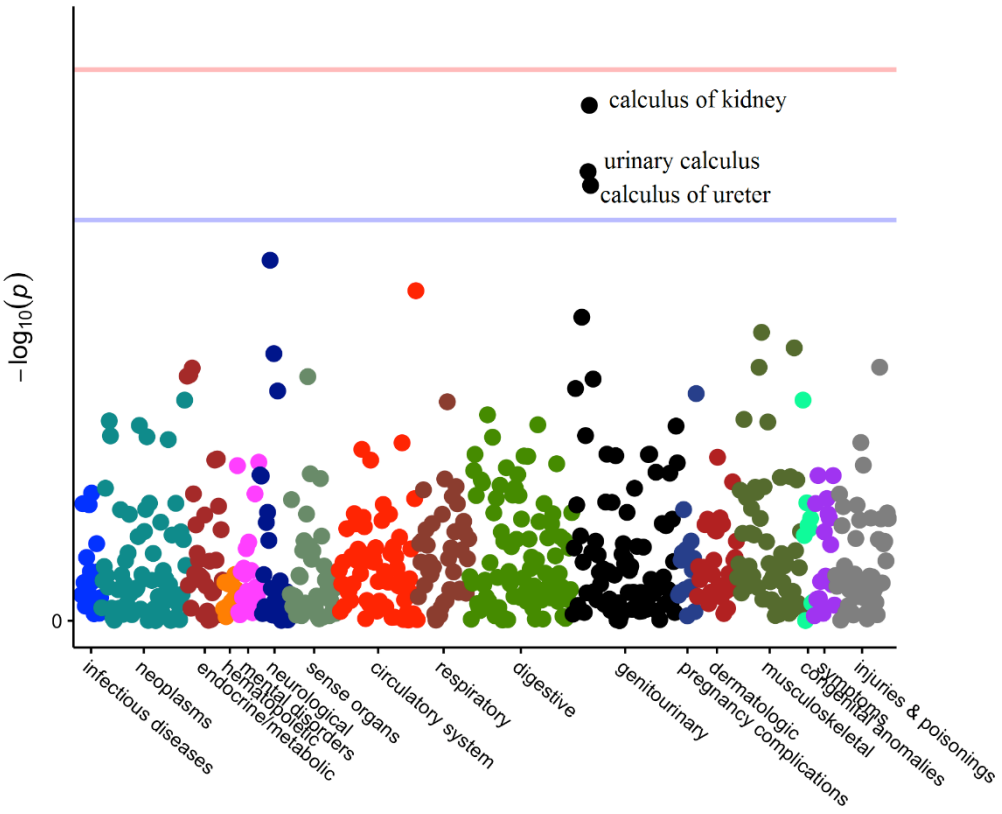
	Continuous		Categorical		
Confounding factors	*F*-value	*P*-value	X^2	*df*	*P*-value
Age	0.154	0.857			
BMI	0.579	0.561			
Time spend outdoors in summer	0.017	0.983			
Time spend outdoors in winter	0.138	0.871			
Assessment center			22.043	42	0.995
Average household income before tax			5.633	12	0.933
Qualification			21.546	14	0.088
Alcohol intake frequency			23.678	12	0.022

Figure 34. Manhattan plot for the PheWAS of rs17216707.



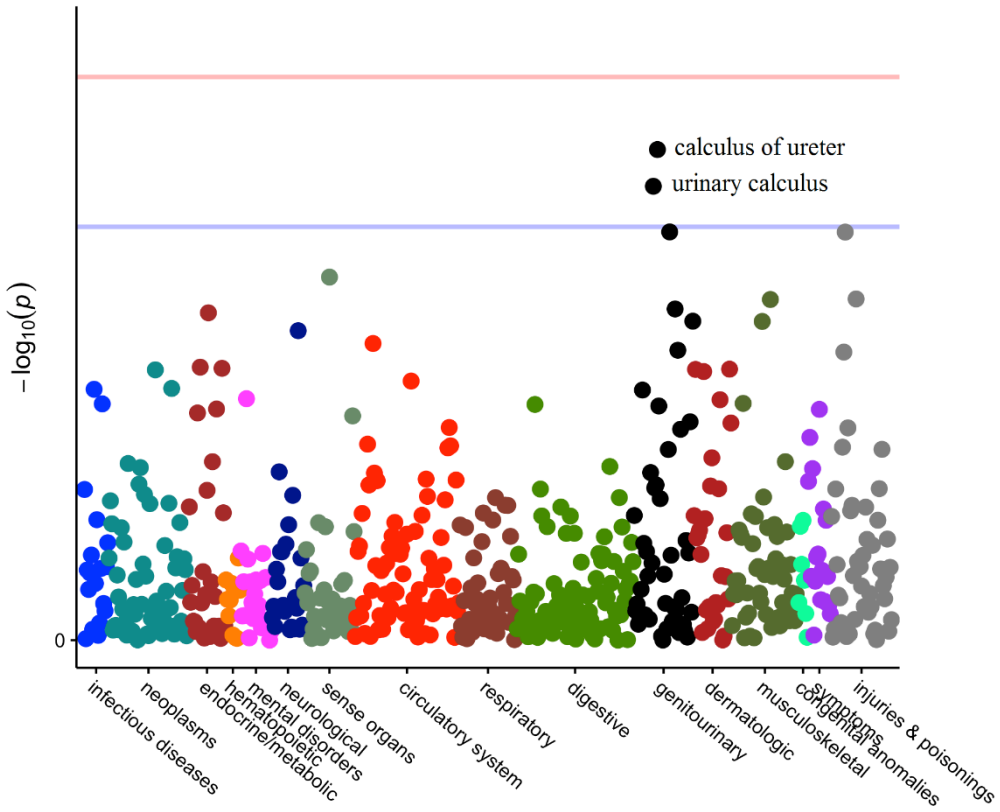
Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $5.44*10^{-5}$, and the blue line indicates a *P* level of 0.001. Three phenotypes survived Bonferroni correction, which were calculus of ureter ($P = 1.14*10^{-6}$), urinary calculus ($P = 1.31*10^{-6}$) and alveolar and parieto-alveolar pneumonopathy ($P = 3.53*10^{-5}$). In addition, there was one phenotype with *P* value less than 0.001, which was calculus of kidney ($P = 2.98*10^{-4}$).

Figure 35. Manhattan plot for the PheWAS of rs17216707 in females.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.35*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There were three phenotypes with *P* value less than 0.001, which were calculus of kidney ($P = 1.39*10^{-4}$), urinary calculus ($P = 4.35*10^{-4}$) and calculus of ureter ($P = 5.51*10^{-4}$).

Figure 36. Manhattan plot for the PheWAS of rs17216707 in males.



Phenotypes aggregated on International Classification of Disease (ICD) codes were plotted with the –log10 *P* value of each association. The red line indicates a Bonferroni corrected *P* level of $7.96*10^{-5}$, and the blue line indicates a *P* level of 0.001. No phenotype survived Bonferroni correction. There were two phenotypes with *P* value less than 0.001, which were calculus of ureter ($P = 2.75*10^{-4}$) and urinary calculus ($P = 5.08*10^{-4}$).

Finally, for a summary of all outcomes which was associated with the score or any SNPs at $P < 0.001$ level and the related P value level comparison, please see **Table 46**, **Table 47** and **Table 48**.

Table 46. Comparison of the *P* values for all outcomes associated with score or SNPs at $P < 0.001$ level.

Outcome	Score	rs3755967	rs10741657	rs12785878	rs10745742	rs8018720	rs17216707
alveolar and parieto-alveolar pneumonopathy	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	Bonferroni
calculus of kidney	<0.05	>0.05	>0.05	>0.05	>0.05	>0.05	<0.001
calculus of ureter	<0.05	>0.05	>0.05	>0.05	>0.05	>0.05	Bonferroni
carcinoma in situ of skin	>0.05	>0.05	<0.05	>0.05	>0.05	>0.05	>0.05
complications of cardiac/vascular device, implant, and graft	>0.05	>0.05	<0.001	>0.05	>0.05	>0.05	>0.05
Delirium	<0.001	<0.05	>0.05	>0.05	>0.05	>0.05	>0.05
labyrinthitis	>0.05	>0.05	<0.001	>0.05	>0.05	>0.05	>0.05
myalgia and myositis	>0.05	>0.05	>0.05	>0.05	>0.05	<0.001	>0.05
nephrotic syndrome	<0.001	>0.05	<0.001	>0.05	>0.05	>0.05	<0.05
Otitis externa	<0.05	<0.05	>0.05	>0.05	>0.05	>0.05	>0.05
pilonidal cyst	<0.05	<0.001	>0.05	>0.05	>0.05	>0.05	>0.05
premature beats	>0.05	>0.05	<0.05	>0.05	>0.05	>0.05	>0.05
rheumatic disease of the heart valves	>0.05	>0.05	>0.05	>0.05	>0.05	<0.001	>0.05
urinary calculus	<0.05	>0.05	>0.05	>0.05	>0.05	>0.05	Bonferroni

Levels of *P* values were shown in the table and shaded. *P* values > 0.05 were shaded with grey; 0.05 to 0.001 were shaded with blue; 0.001 to Bonferroni significance were shaded with yellow; exceeded Bonferroni corrected threshold were shaded with red.

Table 47. Comparison of the *P* values for all outcomes associated with score or SNPs at $P < 0.001$ level in females.

Outcome	Score	rs3755967	rs10741657	rs12785878	rs10745742	rs8018720	rs17216707
alveolar and parieto-alveolar pneumonopathy	NA	NA	NA	NA	NA	NA	NA
calculus of kidney	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	<0.001
calculus of ureter	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	<0.001
carcinoma in situ of skin	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05
complications of cardiac/vascular device, implant and graft	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05
Delirium	NA	NA	NA	NA	NA	NA	NA
labyrinthitis	>0.05	>0.05	<0.05	>0.05	>0.05	<0.05	>0.05
myalgia and myositis	>0.05	>0.05	<0.05	>0.05	>0.05	<0.05	>0.05
nephrotic syndrome	NA	NA	NA	NA	NA	NA	NA
Otitis externa	<0.001	<0.05	>0.05	>0.05	>0.05	>0.05	NA
pilonidal cyst	NA	NA	NA	NA	NA	NA	NA
premature beats	NA	NA	NA	NA	NA	NA	NA
rheumatic disease of the heart valves	>0.05	>0.05	>0.05	>0.05	<0.05	<0.05	>0.05
urinary calculus	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	<0.001

Levels of *P* values were shown in the table and shaded. *P* values > 0.05 were shaded with grey; 0.05 to 0.001 were shaded with blue; 0.001 to Bonferroni significance were shaded with yellow; exceeded Bonferroni corrected threshold were shaded with red.

Table 48. Comparison of the *P* values for all outcomes associated with score or SNPs at $P < 0.001$ level in males.

Outcome	Score	rs3755967	rs10741657	rs12785878	rs10745742	rs8018720	rs17216707
alveolar and parieto-alveolar pneumonopathy	NA	NA	NA	NA	NA	NA	NA
calculus of kidney	<0.05	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05
calculus of ureter	<0.05	>0.05	>0.05	<0.05	>0.05	>0.05	<0.001
carcinoma in situ of skin	>0.05	>0.05	<0.001	>0.05	<0.05	>0.05	>0.05
complications of cardiac/vascular device, implant and graft	>0.05	>0.05	<0.05	>0.05	>0.05	>0.05	>0.05
Delirium	<0.05	<0.05	>0.05	<0.05	>0.05	>0.05	>0.05
labyrinthitis	>0.05	>0.05	<0.05	>0.05	>0.05	>0.05	>0.05
myalgia and myositis	>0.05	>0.05	>0.05	>0.05	>0.05	<0.05	>0.05
nephrotic syndrome	<0.05	>0.05	<0.05	>0.05	>0.05	>0.05	>0.05
Otitis externa	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	NA
pilonidal cyst	<0.05	<0.001	>0.05	>0.05	>0.05	>0.05	<0.05
premature beats	>0.05	>0.05	<0.001	>0.05	>0.05	>0.05	>0.05
rheumatic disease of the heart valves	>0.05	>0.05	>0.05	>0.05	>0.05	<0.05	>0.05
urinary calculus	<0.05	>0.05	>0.05	<0.05	>0.05	>0.05	<0.001

Levels of *P* values were shown in the table and shaded. *P* values > 0.05 were shaded with grey; 0.05 to 0.001 were shaded with blue; 0.001 to Bonferroni significance were shaded with yellow; exceeded Bonferroni corrected threshold were shaded with red.

5.2 Causal association between 25(OH)D level and health outcomes

5.2.1 Power calculation and outcome selection

Table 49 shows statistical power for MR studies of binary outcome under different parameters. If the IV explained 3% of the variance of 25(OH)D level and the case: control ratio was 1:5 or larger, which was always true with such a large cohort as UK Biobank, a binary outcome with 9445 cases or larger would have a power or greater than 80% for detecting a true effect of no less than 1.2 at type I error rate of 0.05. Thus, I considered outcomes with more than 9445 cases as having sufficient statistical power. According to the criteria given in section 4.2.3 (**Figure 14**), a total of 9 outcomes were eligible for subsequent MR analyses (**Table 50**).

Table 49. Case number required in a Mendelian Randomization analysis with a binary outcome with 5% significance level varying effect size (odds ratio) and required statistical power.

Power	OR=1.05	OR=1.10	OR=1.15	OR=1.20	OR=1.25	OR=1.30
70%	103,711	27,177	12,639	7,427	4,958	3,587
75%	116,621	30,561	14,212	8,352	5,575	4,033
80%	131,887	34,561	16,073	9,445	6,305	4,561
85%	150,867	39,535	18,386	10,804	7,213	5,217
90%	176,560	46,268	21,517	12,644	8,441	6,106
95%	218,354	57,220	26,610	15,637	10,439	7,551

Numbers of cases required were calculated assuming 5% significance level, case to control ratio of 1:5, and a 3% explained variance of exposure by instrumental variable, with the equation published by Burgress S. (265).

Table 50. Selective conditions met by the MR included outcomes.

	PheWAS [a]	MR studies [b]	umbrella review [c]	Power [d]
SBP	No	Yes	No	Yes
DBP	No	Yes	No	Yes
Hypertension	No	Yes	Yes	Yes
T2D	No	Yes	Yes	Yes
IHD	No	No	Yes	Yes
BMI	No	No	Yes	Yes

Depression	No	No	Yes	Yes
Non-vertebral fracture	No	No	Yes	Yes
All-cause mortality	No	Yes	No	Yes

Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; T2D, type 2 diabetes; IHD, ischemic heart disease; BMI, body mass index.

[a] Phenotype survived Bonferroni correction in the phenome wide association study for weighted score of 6 SNPs.

[b] Phenotype had significant or conflicting evidences from published Mendelian Randomization studies.

[c] Phenotype was classified as probable or suggestive by the published umbrella review of observational studies and RCTs (25).

[d] Phenotype was of sufficient statistical power.

6.2.2 Incorporation of self-reported medical conditions

I integrated EMR data with self-reported medical condition data for the final MR analysis. The outcomes that I selected were systolic blood pressure (SBP), diastolic blood pressure (DBP), risk of hypertension, risk of type 2 diabetes (T2D), risk of ischemic heart disease (IHD), body mass index (BMI), risk of depression, risk of non-vertebral fracture and all-cause mortality. This merged phenotype dataset increased the number of cases, re-assigned spurious controls, thus increasing the power of our analysis (**Table 51**).

For hypertension, self-reported data included 42317 cases (39.8% of the final case number) which were not captured by the EMR data. The number of hypertension cases was eventually 106,405. For T2D, 671 cases missed by EMR were reported by the self-reported data (4.2%). Incorporating the self-reported data formed a case size of 15,958 for T2D. For IHD, the final case size was 28,337, while 2,566 (9.0%) were missed by EMR but captured by self-reported data. For depression, the final case size was 23,294, a large proportion of which were missed by EMR (13628 cases, 58.5%). For non-vertebral fracture, I finally got 23,603 cases, 6,382 (27.0%) of which were missed by EMR. I got a total of 9830 cases for all-cause mortality, all of which came from the death registry data. SBP, DBP and BMI were measured at baseline recruitment (**Table 51**).

Table 51. Numbers of cases in Mendelian Randomization analysis.

Outcomes	N total [a]	N, EMR [b]	N, SR [c]	N, both [d]
SBP [e]	319778 (100%)	NA	NA	NA
DBP [e]	319779 (100%)	NA	NA	NA
Hypertension	106405 (100%)	16905 (15.9%)	42317 (39.8%)	47183 (44.3%)
T2D	15958 (100%)	13692 (85.8%)	671 (4.2%)	1595 (10.0%)
IHD	28337 (100%)	13062 (46.1%)	2556 (9.0%)	12719 (44.9%)
BMI [e]	338172 (100%)	NA	NA	NA
Depression	23294 (100%)	5382 (23.1%)	13628 (58.5%)	4284 (18.4%)
Non-vertebral fracture	23603 (100%)	15811 (67.0%)	6382 (27.0%)	1410 (6.0%)

All-cause mortality [f]	9830 (100%)	9830 (100%)	NA	NA

Notes: a) Total number of cases; b) Number of cases captured by electronic medical records data only; c) Number of cases captured by self-reported medical data only; d) Number of cases captured by both electronic medical records and self-reported data; e) continuous variable, data came from baseline anthropometric measurement data; f) mortality data, data came from death registry solely.

Abbreviations: EMR, electronic medical records; SR, self-reported medical conditions; SBP, systolic blood pressure; DBP, diastolic blood pressure; T2D, type 2 diabetes; IHD, ischemic heart disease; BMI, body mass index.

6.2.3 Mendelian Randomisation results

I conducted a linear regression of the score on 25(OH)D level in controls from the SOCCS study (n = 2821). The R^2 value was 1.61%, and the *F* statistic was 45.96, indicating that the score is a strong IV for MR (269).

I did not observe any statistically significant associations in our MR analyses for all the tested outcomes and the results from the three different MR methods (two-stage MR, IVW MR and Egger's MR) were consistent (**Table 52, Figure 37**). For continuous outcomes, the effect estimates I got for SBP, DBP and BMI were -0.648 mmHg (se = 0.451, P = 0.210), -0.117 mmHg (se = 0.251, P = 0.661) and 0.130 kg/m^2 (se = 0.121, P = 0.329) per standard deviation increase of the natural log transformed 25(OH)D level. The risk of hypertension decreased by 0.027 (OR = 0.973, 95% CI: 0.911 to 1.040, *P* = 0.340) by every standard deviation increase of the natural log transformed 25(OH)D level. The risk of T2D decreased by 0.029 (OR = 0.971, 95% CI: 0.845 to 1.117, *P* = 0.617) by every standard deviation increase of the natural log transformed 25(OH)D level. The OR for IHD risk was 1.020 (OR = 1.020, 95% CI: 0.917 to 1.135, *P* = 0.647), for depression was 0.913 (OR = 0.913, 95% CI: 0.816 to 1.022, P = 0.093), for non-vertebral fracture was 0.969 (OR = 0.969, 95% CI: 0.867 to 1.083, *P* = 0.495) and for all-cause mortality was 1.030 (OR = 1.030, 95% CI: 0.869 to 1.222, *P* = 0.671) by every standard deviation increase of the natural log transformed 25(OH)D level. For all the above outcomes, due to the substantial power, I reached effect estimates close to null with very narrow confidence interval (i.e., high precision of the estimates).

Additionally, I implemented the Egger's MR. The intercept item of Egger's MR tests for the existence of unbalanced pleiotropy. If there are no pleiotropic effects among all the instruments, or there are pleiotropic effects of several instruments in opposite directions and they cancelled out, the intercept item would not be statistically significant ($P > 0.05$). This means that the effect estimates derived from two-stage MR and IVW MR are valid and precise. In my MR Egger's regressions, the *P* values of the intercept term for all outcomes were greater than 0.05, indicating no evidence of unbalanced pleiotropy among the variants I used (**Table 53**).

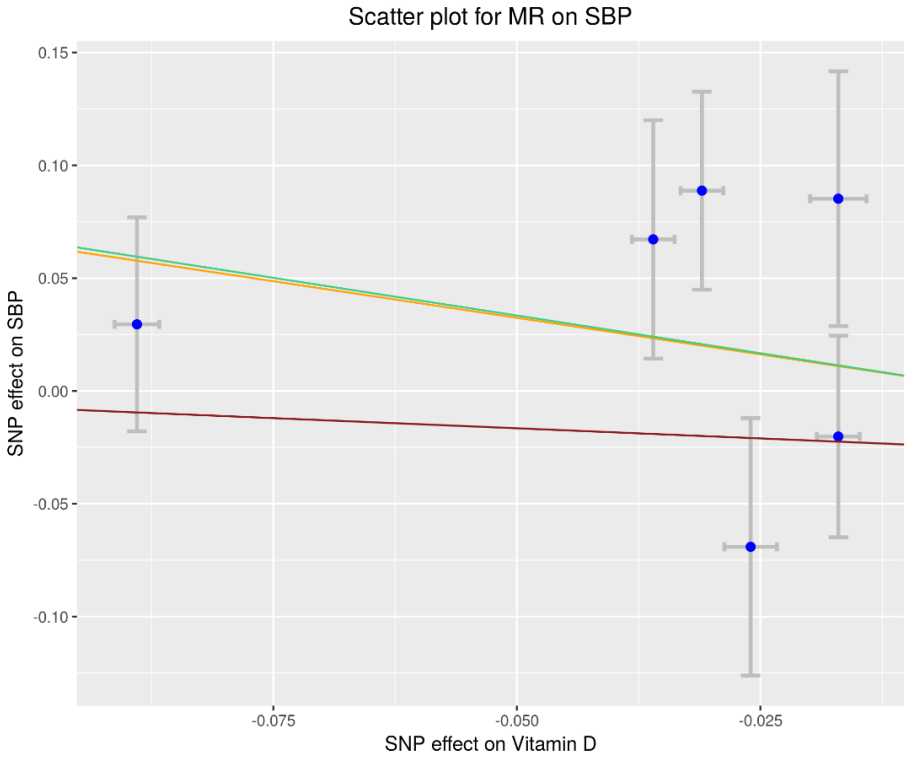
Table 52. Mendelian Randomization causal effect estimates.

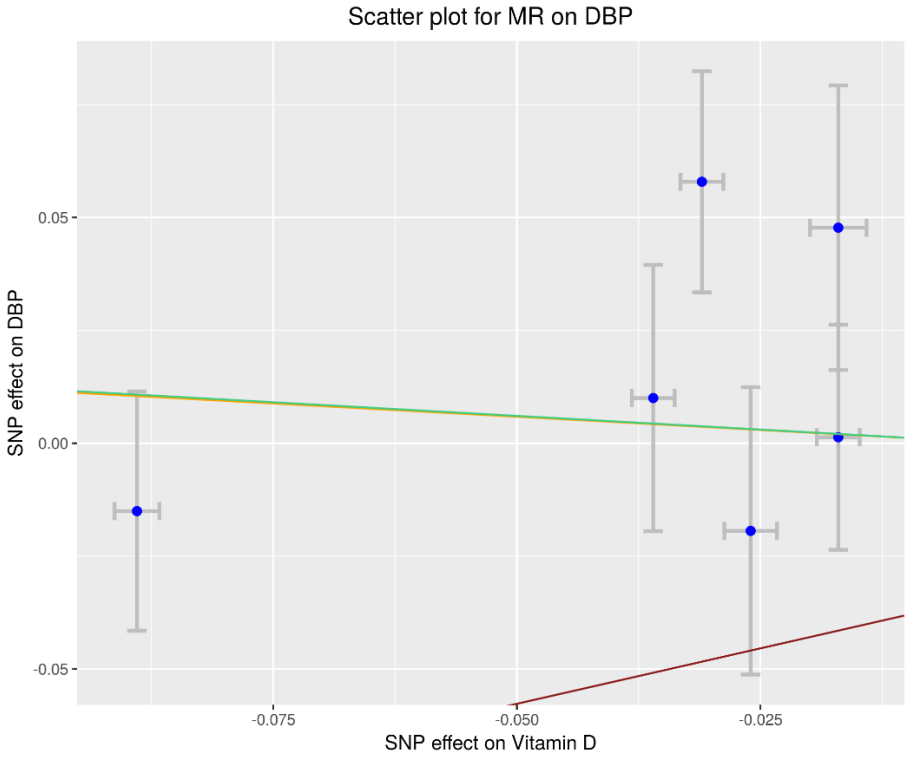
Method	beta	se	*P*-value	OR	95% CI	N total/N cases	Power
Systolic blood pressure						319778	NA
two-stage MR	-0.669	0.449	0.137	NA	NA		
IVW MR	-0.648	0.451	0.210	NA	NA		
Egger's regression	-0.180	1.086	0.876	NA	NA		
Diastolic blood pressure						319779	NA
two-stage MR	-0.121	0.251	0.629	NA	NA		
IVW MR	-0.117	0.251	0.661	NA	NA		
Egger's regression	0.491	0.530	0.407	NA	NA		
Hypertension						339256/106405	1.00
two-stage MR	-0.056	0.059	0.343	0.976	0.928-1.026		
IVW MR	-0.063	0.060	0.340	0.973	0.911-1.040		
Egger's regression	0.084	0.175	0.657	1.037	0.841-1.278		
Type 2 Diabetes						339256/15958	0.97
two-stage MR	-0.060	0.126	0.632	0.974	0.876-1.083		
IVW MR	-0.067	0.126	0.617	0.971	0.845-1.117		
Egger's regression	0.242	0.244	0.377	1.110	0.829-1.485		
Ischaemic Heart Disease						339256/28337	1.00
two-stage MR	0.049	0.096	0.611	1.021	0.942-1.107		
IVW MR	0.047	0.096	0.647	1.020	0.917-1.135		
Egger's regression	0.109	0.219	0.645	1.048	0.807-1.360		
Body mass index						338172	NA
two-stage MR	0.128	0.120	0.288	NA	NA		
IVW MR	0.130	0.121	0.329	NA	NA		
Egger's regression	-0.099	0.213	0.665	NA	NA		
Depression						339256/23294	1.00
two-stage MR	-0.216	0.102	0.034	0.911	0.837-0.993		
IVW MR	-0.212	0.102	0.093	0.913	0.816-1.022		
Egger's regression	-0.311	0.180	0.158	0.875	0.706-1.084		
Non-vertebral fracture						339256/23603	1.00
two-stage MR	-0.068	0.101	0.497	0.971	0.892-1.057		
IVW MR	-0.074	0.101	0.495	0.969	0.867-1.083		
Egger's regression	-0.092	0.265	0.747	0.961	0.700-1.320		
All-cause mortality						339256/9830	0.87
two-stage MR	0.073	0.154	0.634	1.032	0.907-1.175		
IVW MR	0.069	0.154	0.671	1.030	0.869-1.222		
Egger's regression	0.192	0.272	0.520	1.086	0.785-1.503		

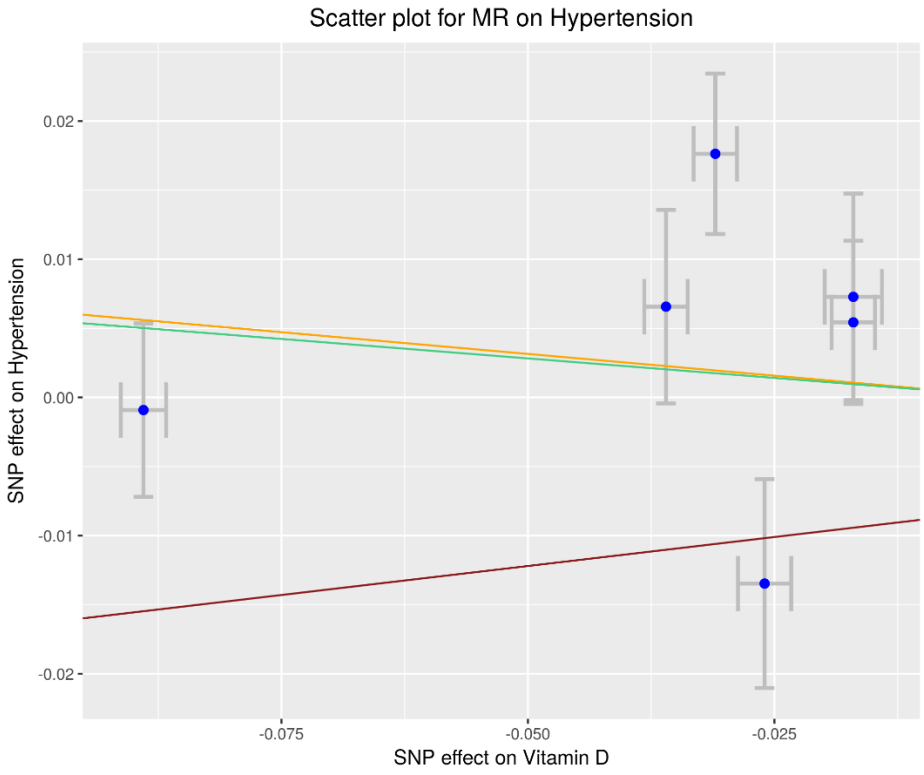
Abbreviations: MR, Mendelian Randomisation; IVW, inverse variance weighted; OR, odds ratio.

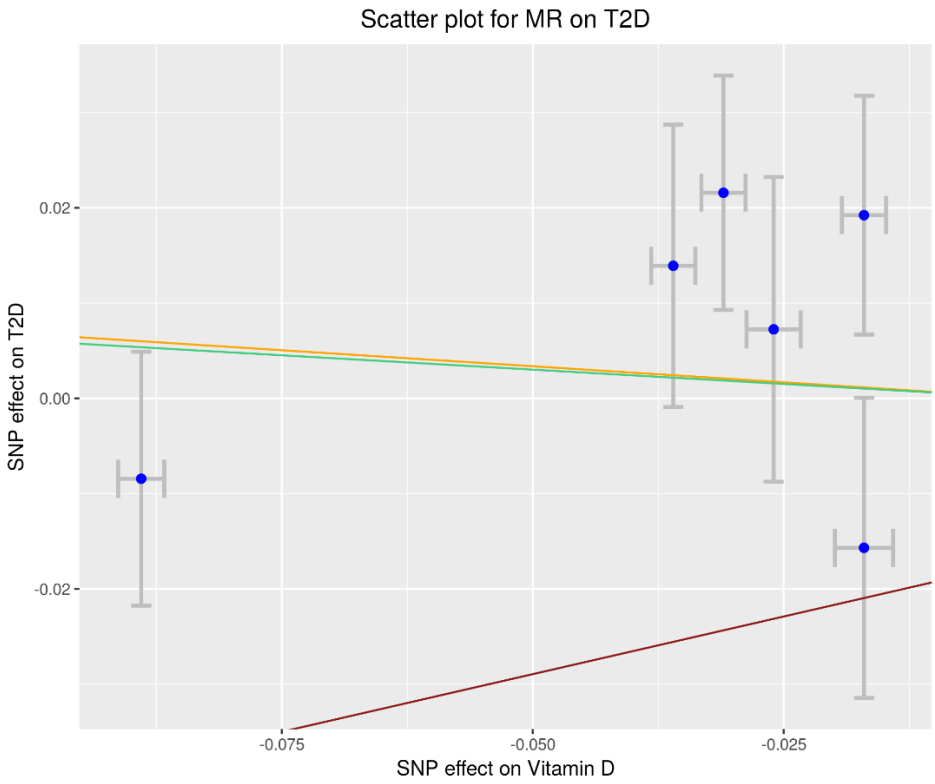
Notes: MR effect estimates were done with three different MR methods. OR was calculated as exponential of beta*sd (the standard deviation of log transformed 25(OH)D level in an independent British population, SOCCS, which was 0.430), whose unit was per sd increase of log transformed 25(OH)D level. The upper/lower 95% CI was calculated similarly; with the same unit as OR. Power was calculated assuming a R square of 0.03, OR of 1.2 and significance level at 0.05.

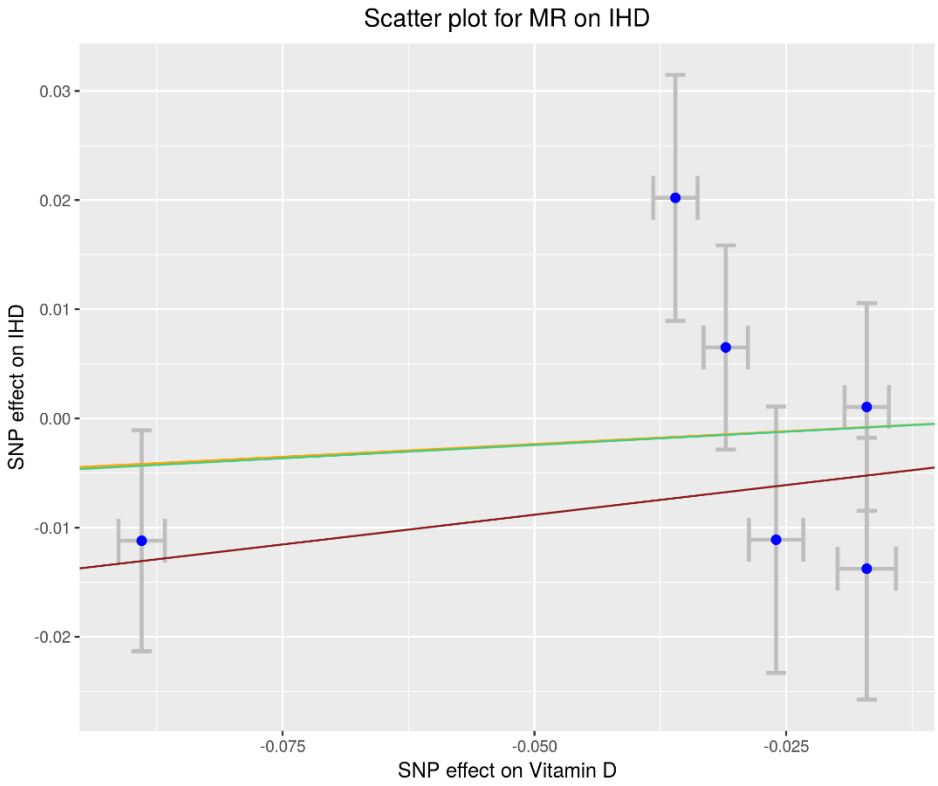
Figure 37. Scatter plots for Mendelian Randomization on nine outcomes

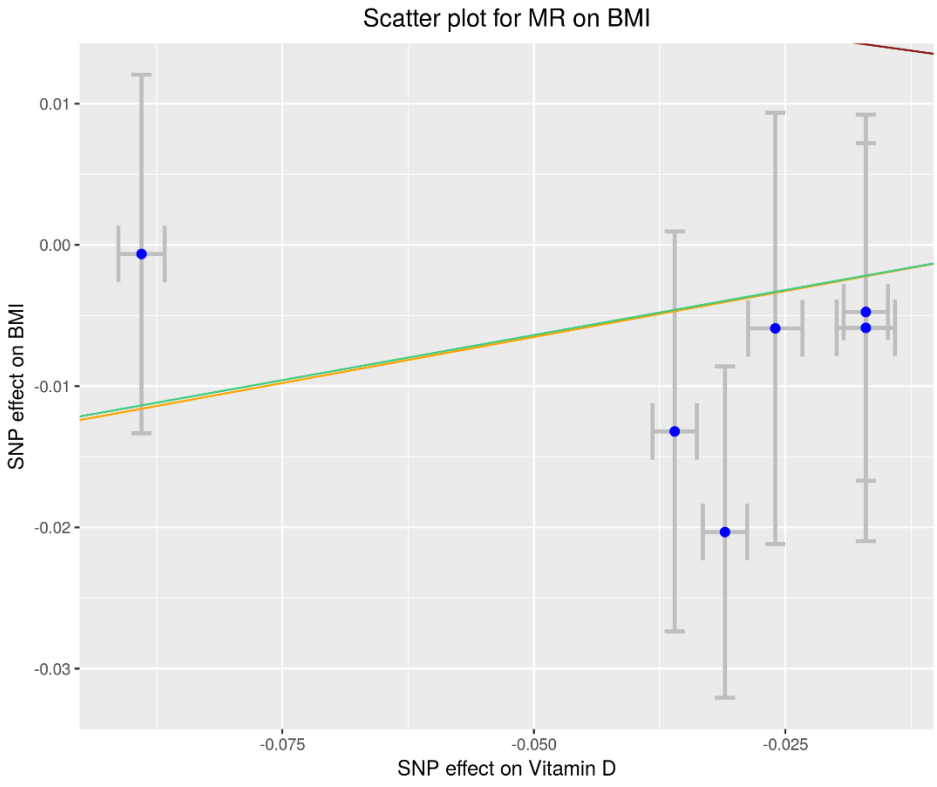


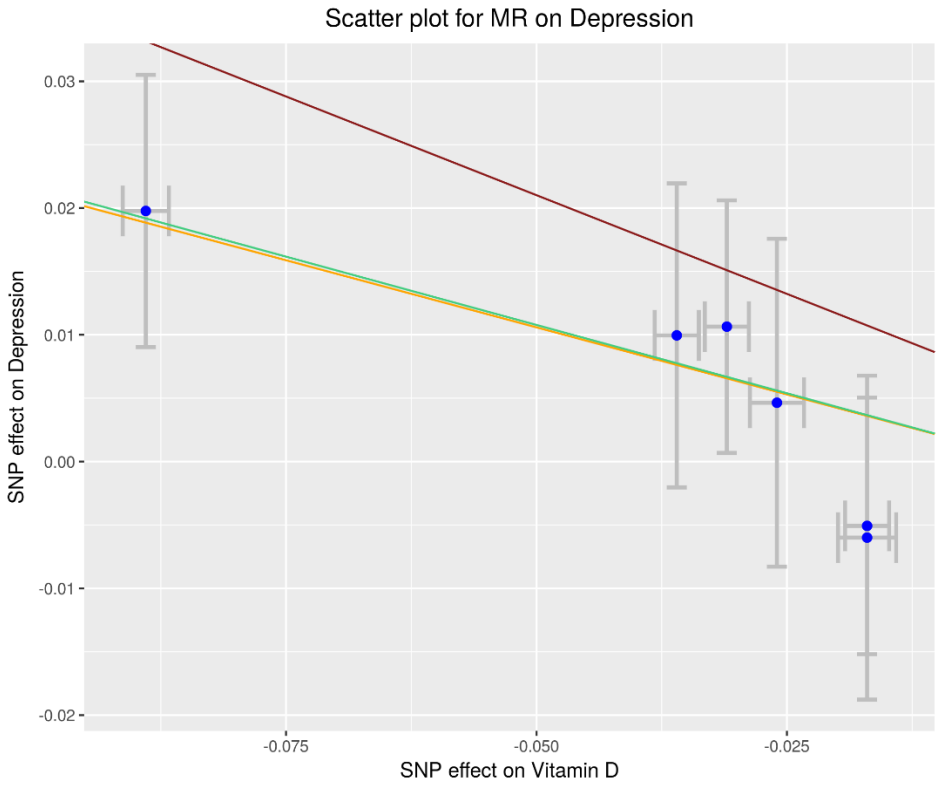


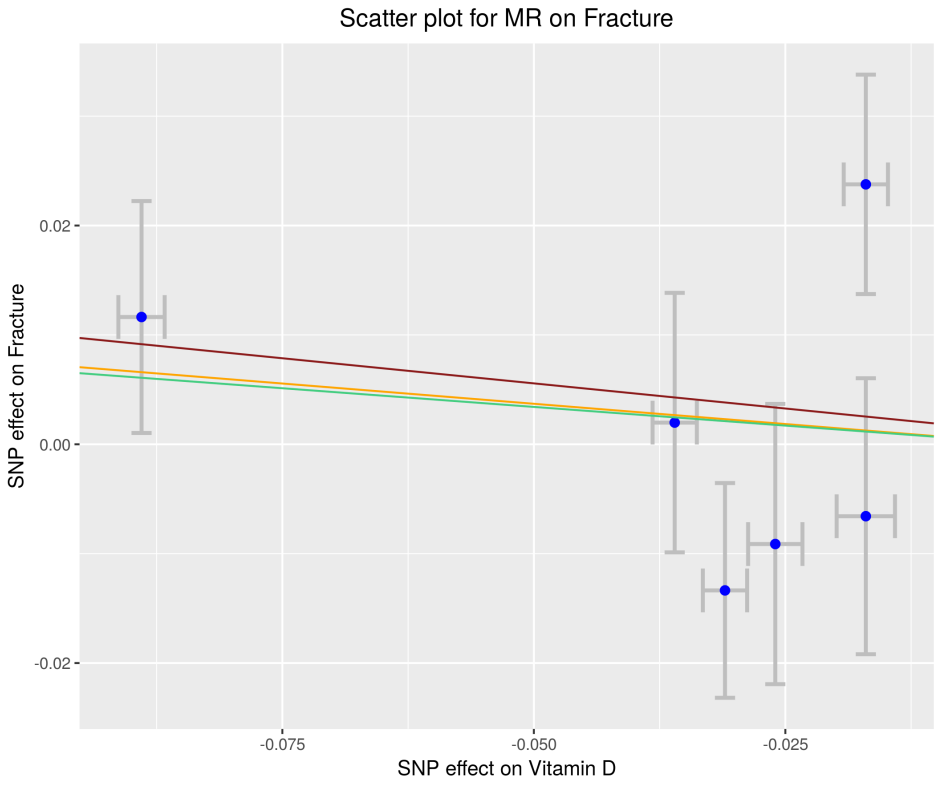

Scatter plot for MR on Hypertension
SNP effect on Hypertension
0.02
0.01
0.00
-0.01
-0.02
-0.075
-0.050
-0.025
SNP effect on Vitamin D

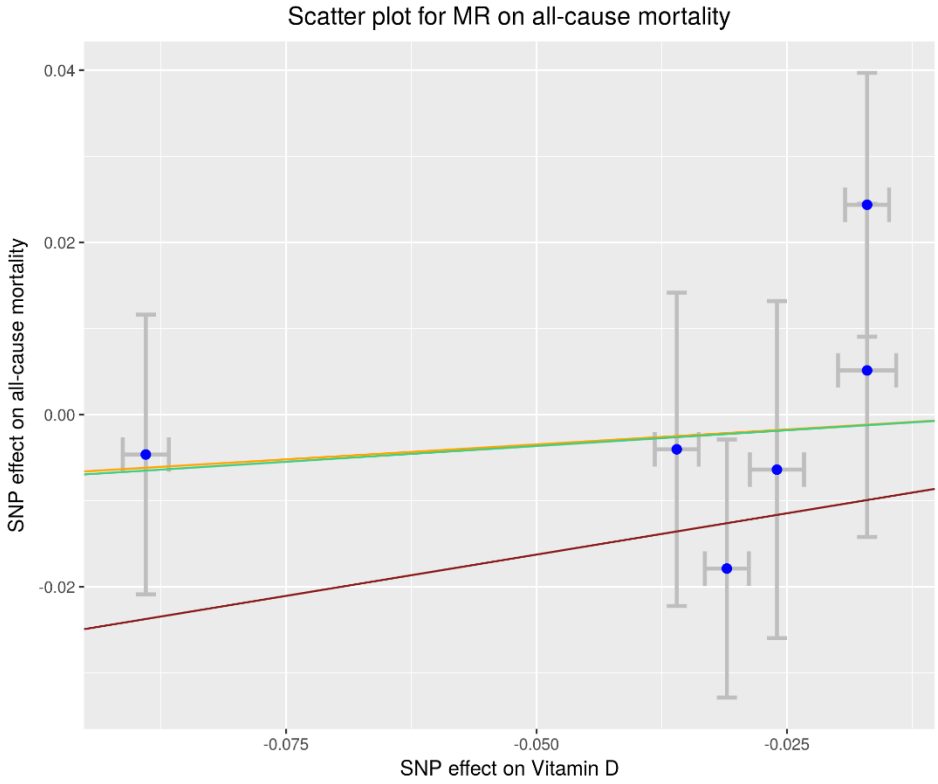

Scatter plot for MR on T2D
SNP effect on T2D
0.02
0.00
-0.02
-0.075
-0.050
-0.025
SNP effect on Vitamin D


Scatter plot for MR on IHD
SNP effect on IHD
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.075
-0.050
-0.025
SNP effect on Vitamin D


Scatter plot for MR on BMI
SNP effect on BMI
0.01
0.00
-0.01
-0.02
-0.03
-0.075
-0.050
-0.025
SNP effect on Vitamin D


Scatter plot for MR on Depression
SNP effect on Depression
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.075
-0.050
-0.025
SNP effect on Vitamin D


Scatter plot for MR on Fracture
SNP effect on Fracture
0.02
0.00
-0.02
-0.075
-0.050
-0.025
SNP effect on Vitamin D



X axis represents SNPs' effects on 25(OH)D level. Y axis represents SNPs' effects on outcomes. The green line represents the regression line from inverse variance weighted Mendelian Randomization. The yellow line represents the regression line from two-stage Mendelian Randomization. The red line represents the regression line from Egger's Mendelian Randomization.

Table 53. Results for intercept items of Egger's regressions.

Phenotype	beta	s.e.	*P* value
SBP	-0.026	0.049	0.628
DBP	-0.033	0.024	0.236
Hypertension	-0.008	0.008	0.366
T2D	-0.017	0.011	0.198
IHD	-0.003	0.010	0.748
BMI	0.013	0.010	0.260
Depression	0.019	0.012	0.203
Non-vertebral fracture	0.002	0.011	0.879
All-cause mortality	-0.007	0.012	0.614

Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; T2D, type 2 diabetes; IHD, ischemic heart disease; BMI, body mass index; s.e., standard error.

Beta, standard error and *P* values for intercept items of Egger's regression for every outcome. Significance ($P<0.05$) of the intercept item indicates existence of unbalanced pleiotropy.

6.2.4 Sensitivity analyses

Leave-one-out analyses

I conducted leave-one-out sensitivity analyses for the 9 selected outcomes with the IVW MR. For each single outcome, I dropped one SNP at a time, and conducted IVW MR involving the other five SNPs as IVs. For all the nine outcomes, leaving any SNP out did not change the results, which indicates that the results were not due to effects of a specific SNP (**Table 54** to **Table 62**).

Table 54. Leave-one-out sensitivity analysis for systolic blood pressure.

SNP	beta	se	*P*-value
rs3755967	-1.440	0.843	0.163
rs10741657	-0.399	0.475	0.449
rs12785878	-0.522	0.473	0.332
rs10745742	-0.704	0.457	0.199
rs8018720	-0.567	0.455	0.281
rs17216707	-0.794	0.460	0.160

Table 55. Leave-one-out sensitivity analysis for diastolic blood pressure.

SNP	beta	se	*P*-value
rs3755967	-0.833	0.471	0.151
rs10741657	0.0803	0.265	0.777
rs12785878	-0.101	0.264	0.723
rs10745742	-0.118	0.255	0.667
rs8018720	-0.067	0.254	0.805
rs17216707	-0.155	0.257	0.578

Table 56. Leave-one-out sensitivity analysis for risk of hypertension.

SNP	beta	se	*P*-value	OR	95% CI
rs3755967	-0.246	0.112	0.092	0.782	0.573-1.066
rs10741657	-0.006	0.063	0.930	0.994	0.835-1.184
rs12785878	-0.051	0.063	0.465	0.951	0.799-1.131
rs10745742	-0.055	0.061	0.414	0.946	0.800-1.120

rs8018720	-0.056	0.060	0.404	0.945	0.800-1.117
rs17216707	-0.088	0.061	0.220	0.915	0.773-1.084

Table 57. Leave-one-out sensitivity analysis for risk of type 2 diabetes.

SNP	beta	se	*P*-value	OR	95% CI
rs3755967	-0.472	0.236	0.117	0.624	0.324-1.203
rs10741657	0.004	0.133	0.979	1.004	0.693-1.454
rs12785878	-0.034	0.133	0.810	0.967	0.668-1.398
rs10745742	-0.035	0.128	0.798	0.965	0.676-1.379
rs8018720	-0.086	0.128	0.537	0.917	0.644-1.308
rs17216707	-0.058	0.129	0.677	0.944	0.659-1.351

Table 58. Leave-one-out sensitivity analysis for risk of ischemic heart disease.

SNP	beta	se	*P*-value	OR	95% CI
rs3755967	-0.151	0.180	0.449	0.860	0.522-1.417
rs10741657	0.076	0.101	0.496	1.079	0.814-1.430
rs12785878	0.110	0.101	0.337	1.116	0.843-1.478
rs10745742	0.050	0.098	0.635	1.051	0.802-1.378
rs8018720	0.032	0.097	0.756	1.033	0.789-1.352
rs17216707	0.030	0.098	0.774	1.031	0.785-1.354

Table 59. Leave-one-out sensitivity analysis for body mass index.

SNP	beta	se	*P*-value
rs3755967	0.438	0.226	0.124
rs10741657	0.071	0.127	0.607
rs12785878	0.106	0.127	0.451
rs10745742	0.126	0.122	0.362
rs8018720	0.126	0.122	0.358
rs17216707	0.126	0.123	0.364

Table 60. Leave-one-out sensitivity analysis for risk of depression.

SNP	beta	se	*P*-value	OR	95% CI
rs3755967	-0.186	0.191	0.387	0.831	0.488-1.413

rs10741657	-0.197	0.108	0.141	0.821	0.609-1.107
rs12785878	-0.205	0.107	0.128	0.815	0.605-1.097
rs10745742	-0.227	0.104	0.093	0.797	0.598-1.062
rs8018720	-0.222	0.103	0.097	0.801	0.602-1.066
rs17216707	-0.213	0.104	0.110	0.808	0.605-1.079

Table 61. Leave-one-out sensitivity analysis for risk of non-vertebral fracture.

SNP	beta	se	*P*-value	OR	95% CI
rs3755967	0.069	0.189	0.735	1.071	0.634-1.810
rs10741657	-0.131	0.106	0.285	0.877	0.653-1.178
rs12785878	-0.076	0.106	0.512	0.927	0.691-1.243
rs10745742	-0.034	0.102	0.754	0.966	0.727-1.284
rs8018720	-0.083	0.102	0.462	0.921	0.694-1.221
rs17216707	-0.093	0.103	0.419	0.912	0.685-1.213

Table 62. Leave-on-out sensitivity analysis for risk of all-cause mortality.

SNP	beta	se	*P*-value	OR	95% CI
rs3755967	0.113	0.289	0.716	1.119	0.502-2.498
rs10741657	0.012	0.163	0.945	1.012	0.644-1.590
rs12785878	0.065	0.162	0.709	1.067	0.680-1.674
rs10745742	0.115	0.157	0.504	1.122	0.726-1.733
rs8018720	0.076	0.156	0.649	1.079	0.700-1.663
rs17216707	0.062	0.158	0.715	1.064	0.687-1.648

Analyses excluding variants associated with UK Biobank assessment centres

Since three SNPs (rs10741657, rs12785878 and rs10745742) were significantly associated with UK Biobank assessment centre I implemented a sensitivity analysis with the IVW MR, only including the other 3 SNPs (rs3755967, rs8018720 and rs17216707).

Similar to the results of the MR of all 6 SNPs, I did not find any statistically significant association in this sensitivity analysis. All outcomes had effect estimates similar with the 6-SNP MR, but wider confidence intervals (**Table 63**).

Table 63. Inverse variance weighted Mendelian Randomization with 3 SNPs.

Phenotype	beta	se	*P*-value	OR	95% CI
SBP	-0.280	0.512	0.639	NA	NA
DBP	0.130	0.286	0.694	NA	NA
Hypertension	0.027	0.068	0.724	1.012	0.893-1.147
T2D	0.094	0.144	0.579	1.041	0.798-1.359
IHD	0.159	0.109	0.284	1.071	0.875-1.310
BMI	0.027	0.137	0.861	NA	NA
Depression	-0.206	0.116	0.217	0.915	0.739-1.134
Non-vertebral fracture	-0.093	0.114	0.504	0.961	0.778-1.188
All-cause mortality	0.054	0.175	0.786	1.024	0.740-1.416

Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; T2D, type 2 diabetes; IHD, ischemic heart disease; BMI, body mass index; s.e., standard error.

Analyses were done with inverse variance weighted MR, using rs3755967, rs8018720 and rs17216707 as instrumental variables. OR was calculated as exponential of beta*sd (the standard deviation of log transformed 25(OH)D level in an independent British population, SOCCS, which was 0.430), whose unit was per sd increase of log transformed 25(OH)D level. The upper/lower 95% CI was calculated similarly; with the same unit as OR.

Chapter VI DISCUSSION

6.1 Introduction

In the first chapter of this thesis, multiple aspects of vitamin D were introduced. I first gave a brief introduction on vitamin D, its synthesis and metabolism, its biological functions and the severity of vitamin D deficiency and insufficiency worldwide. Then, I updated an umbrella review, which summarized evidence on vitamin D and health outcomes from previously published meta-analyses of observational studies and RCTs, by searching for articles published between 2014 and 2018. In addition, I searched PubMed and the GWAS catalogue for all published GWAS on 25-hydroxyvitamin D 25(OH)D concentration and summarized all the loci/genes which have been found to be associated with 25(OH)D. Finally, I conducted a systematic literature review for all previously published Mendelian Randomization studies on association between 25(OH)D level and health outcomes, which had not been covered by the umbrella review of meta-analyses of observational studies and RCTs. In chapter two, I presented a systematic literature review on published PheWAS studies and summarized the methodology that was applied and the main findings from the included studies. I also presented a general pipeline on how to conduct and report PheWAS studies and discussed its advantages, disadvantages and future developments. In chapter three, I presented the aims and objectives and in chapter four, I described the UK Biobank cohort and the specific methods I used. All relevant results were reported in chapter five.

In this chapter, I will discuss the main issues, which emerged in this study. In the first part of the discussion, issues regarding the methodological and analytical aspects of this study are presented. In particular, I discuss the strengths and limitations of the study design, population, and statistical methods I used. In the second part of this chapter, the most important findings from this study are discussed and compared with findings from previous published studies.

6.2 Methodological and analytical issues

In this part of the chapter, I will first discuss the strengths and limitations of the UK Biobank population. Then strengths and limitations of the PheWAS method and the

MR method, which arose in my thesis, will be discussed.

6.2.1 The UK Biobank cohort

6.2.1.1 Strengths of the UK Biobank

The large sample size

As described in chapter IV, the UK Biobank is a British cohort that recruited participants across the UK. With a sample size of ~500,000, the UK Biobank provides a very big population for studies on various topics. For outcomes of large sample size the statistical power is improved, which can be a significant challenge for a MR study (265). As presented in **Table 49**, I had an estimated statistical power of 90% for outcomes with more than 46,268 cases at OR of 1.10, and for outcomes with more than 12,644 cases at OR of 1.20. Therefore, in this cohort, there should be sufficient power for detecting moderate to large effect sizes for common diseases (e.g., hypertension, cardiovascular diseases). In contrast, eleven out of the 27 identified MR studies in the systematic literature review (**Table 7**, 11/27 = 40.7%) were of sample size less than 10,000 (182, 184, 185, 188, 190, 192, 193, 196, 270-272).

This also holds true for the phenome wide analysis, which used logistic regression. From my systematic literature review for PheWAS studies, 24 out of the 45 PheWAS studies (**Table 8**, 10/20 = 50.0%) were of sample size less than 10,000 (Table 8) (199, 201, 203-205, 207, 210, 216, 236, 237). However, since large number of phenotypes are tested, multiple testing burden is substantial for PheWAS. A PheWAS of 10,000 individuals is not sufficient for surviving the multiple testing correction, especially for moderate to small effect outcomes. Among the studies with sample size greater than 10,000, only five of them were of sample size larger than 100,000, ranging from 521,000 to 1,749,400 (212, 217, 226, 228, 241). In addition, although hundreds or thousands of distinct phenotypes could be defined through linkage with EMR, many PheWAS only analyses phenotypes with more than 20 or 25 cases (due to power consideration). Consequently, the number of phenotypes which are selected to go to statistical analysis will be small for PheWAS study of limited sample size.

Therefore, this study had larger statistical power for common outcomes of moderate

to large effects compared to previous PheWAS and MR studies of vitamin D.

Comprehensive data collection

The UK Biobank collected a wide variety of variables, covering demographic factors, environmental exposures, psychosocial factors, environmental factors, lifestyle information, health status and physical and functional measurements (e.g., blood pressure, cognitive functions and hearing threshold). In addition, the UK Biobank also acquired hospital inpatient data, cancer and death registry data and asked for the health status and medical history in baseline assessment and verbal interviews with participants. This expanded the number of phenotypes, which could be tested in a PheWAS or MR, making my study even more comprehensive than previous studies.

In addition, most previous studies used only EMR data in defining case-control status, which can lead to misclassification (will be further discussed in 6.2.1.2). By incorporating the self-reported medical data, medication history and serum/urine biomarkers, the UK Biobank cohort allows a better definition for cases and controls.

Relative homogeneity of UK Biobank study population

Participants of UK Biobank were recruited exclusively in the UK. They were recruited, measured, and interviewed according to the same protocol in all 22 assessment centres. Moreover, I restricted my analysis in white British participants based on their self-reported ethnic background and ancestral PCs calculated from their genetic data. Therefore, the set of participants I studied was homogenous in recruitment, genotype/phenotype measurement, and genetic ancestry. This is expected to be advantageous since it is known that heterogeneity and population stratification cause bias and affect the implications of epidemiological studies, as discussed below.

First, humans vary at both geographical and individual level (273). Populations of different races may vary in their physiological pathways and associations which exist in one ethnic group are not necessarily true in another (i.e., population differentiation). For example, in a study of Chinese participants, which explored the associations between BMI-related loci identified by GWAS from white population and BMI or

obesity in Chinese population, few of the loci were successfully replicated. From their allele frequency calculations, evidence of population stratification was shown between Chinese individuals and Europeans (274). Specifically for my study, the associations between vitamin D and health outcomes may not be universally identical for Europeans residing in different areas. Since European-ancestral individuals residing in different countries have greatly differing sun exposures, they can be expected to have different 25(OH)D levels. At different 25(OH)D levels, associations dominant in countries of lower latitude may not exist in the UK population. For example, I failed to replicate observed causal associations between 25(OH)D and MS (180, 181). Although the results from my study could be criticised for lack of power for the MS outcome, since there were only 986 cases, I also noticed that the inferred OR reported by a previous MR study was 2.0 (95%CI: 1.7-2.5, $P = 7.7*10^{-12}$) (180), which was a large effect. With a case number of 986 and a true OR of 2.0, the estimated power for my analyses was nearly 100%. However, it should be acknowledged that the OR of 2.0 originally reported by Mokry and colleagues was vulnerable to winner's curse, where newly discovered true associations often have inflated effects compared to true effects (275). This inflation may be caused by crossing predefined *P* threshold in an underpowered sample, flexible analyses and selective reporting, which was not investigated in the original paper by Mokry and colleagues (180, 275). Thus, the difference between my result for MR and previous studies could be possibly due to actual biological differences between UK population and other white populations in lower latitude. The findings and the reason underlying them need to be investigated by other large studies of UK population.

Secondly, in a meta-analysis of different studies, which is similar to a crude pooled analysis for participants from different ancestral backgrounds and geographically locations, it is essential to check that effect estimates from individual studies are similar enough (i.e., homogeneous). In a meta-analysis, the presence of heterogeneity may affect the statistical validity of the summary estimated effect (276). We can test whether there is more variation between studies than would be expected by chance alone. Rejection of the null hypothesis indicates the existence of heterogeneity. Heterogeneity could be quantified by I^2, which represents the fraction of the total

variance in the meta-analysis estimate that is due to intrinsic variability in the effect size, as distinct from variability arising from measurement error (277). When there is relative homogeneity a fixed-effect model is used in meta-analysis. Alternatively, when heterogeneity exists, a random-effects meta-analysis should be conducted, which takes the between-study variation into consideration (278). However, a valid random-effects model depends on the assumption that effect estimates from individual studies follow a certain distribution pattern, such as a normal distribution. Establishing the validity of distributional assumptions is hard, especially for meta-analysis of small number of studies (279). Therefore, compared to meta-analysis or consortia with a similar pooled sample size, the UK Biobank provides a large prospective cohort of a relatively homogeneous population, which is less vulnerable to problems caused by heterogeneity. Conversely, it should be acknowledged that homogeneity can also limit statistical power, because there may be less population variability in phenotype values. For example it is expected that 25(OH)D levels of UK Biobank participants would vary less compared to the global population.

There are several previous PheWAS and MR studies with sample sizes larger than 100,000, which were comparable to the UK Biobank cohort in size but based on consortia data (179, 180, 186, 195, 279). For instance, the MR study on multiple sclerosis published by Mokry et al. had a total sample size of 14,498 multiple sclerosis cases and 24,091 controls from the International Multiple Sclerosis Genetics Consortium study (IMSGC). The MS cases of IMSGC were collected in 12 countries across the world, including Australia, New Zealand, Belgium, Denmark, Finland, France, Germany, Italy, Norway, Sweden, UK and the US, and cases from different sites varied in female/male rate and mean age of disease onset (280). Large study populations generated by pooling samples from different studies within research group consortia have two potential problems: differences in phenotype measurements, and heterogeneity/population stratification. Each study has its own study protocol, in most cases, and thus measure and record different phenotypes. Meanwhile, the same phenotype is measured by different methods (outcomes defined according to different criteria) and employ different quality control procedures. Thus, harmonisation of phenotypes from different individual studies can be problematic. Phenotype

heterogeneity, in both outcomes and covariates across studies, represents a major challenge to successful meta-analysis of common traits (281). If stark difference exists in phenotype definition, an effort at harmonization is essential (282). Although tools such as PhenX Toolkit of standardized, high priority measures is available to investigators planning new studies, most current consortia involve existing studies whose phenotypes and data collection instruments are already defined (281). This is the reason why some previous PheWAS manually created phenotype classes and categorized the various phenotypes from individual study sites into predefined classes (233, 234, 236). Even though they can be harmonized, this process of binning lost some information (234). In addition, in the situation where samples were collected from broad geographical areas, population stratification can be a problem as has been discussed above. Especially in MR analyses, population stratification violates the IV assumption, and it can bias the results. Therefore, studying associations between exposure and health outcome in a world-wide population should be implemented and interpreted carefully, and the issue of heterogeneity between population of individual studies and population stratification should be considered and tested. In their MS paper, Mokry and colleagues did test for heterogeneity and found an increased I^2 (I^2 = 63%, 95% CI: 0% to 88%). Therefore, they compared the fixed effect model estimates with the random-effects model estimates. However, they did not test whether effect estimates from different centres complied with the random-effects model assumption. In addition, since they used only 4 SNPs as their IVs, they claimed that the heterogeneity of their study could not be accurately measured by I^2 only (180).

6.2.1.2 Limitations of the UK Biobank

Misclassification of case/control status

The phenotyping method I applied may have misclassified the case/control status. The UK Biobank has only released inpatient, death registry and cancer registry data, and therefore my phenotyping was based on those three data sources. However, in the NHS health care system, diseases which are less life threatening (e.g., non-vertebral fracture, depression, hypertension, vitamin D deficiency) are usually not monitored in hospitals, and thus will not be captured by the EMR data. Thus, the precision of phenotyping varies by phenotype. For instance, cases of depression and non-vertebral fracture were

mistakenly classified as controls as be seen from Table 51. A total of 13628 cases for depression (58.5% of the total case number in EMR plus self-reported data) and 6382 cases for non-vertebral fracture (27.0% of the total case number in EMR plus self-reported data) were not captured by the inpatient EMR data. Even for outcomes which were normally monitored in the inpatient department, cases occurred outside of the NHS-funded hospitals would not be captured by the UK Biobank data and thus may also be misclassified as controls.

Although UK Biobank has the self-reported medical data collected from the baseline questionnaire and a verbal interview with a trained research nurse, it is impossible to use this data in PheWAS. The coding of the self-reported health outcomes followed a customized tree-structure system defined by the UK Biobank, which was different from the ICD coding used by EMR data. Due to this difference of coding structures, automatic merging of the EMR data and self-reported data for all outcomes is not feasible with standard computation tools. In the MR analysis of nine outcomes I manually merged those two sources of data. This substantially increased the number of cases for some outcomes, as shown in the results section (Table 51). For example, the merged data identified 106,405 cases for hypertension (42,317 (39.8%) from self-reported data), 23,294 cases for depression (13,628 (58.5%) from self-reported data) and 23,603 for non-vertebral fracture (6,382 (27.0%) from self-reported data). Other outcomes, like type 2 diabetes and ischemic heart disease, were captured more accurately by EMR data and only 4.2% and 9.0% of the total number of cases were exclusively captured by self-reported data.

However, self-reported medical conditions are blamed for their accuracy. In a systematic literature review by the UK Biobank group, which included 17 studies comparing patient self-reported stroke against reference standard, authors found that the positive predictive values of self-reported stroke varied from 22% to 87%. The positive predictive value increased with stroke prevalence. Hence, they concluded that in population-based studies, such as the UK Biobank, a large proportion of self-reported strokes may be false positives (283). In a study of self-reported data accuracy for fractures in elderly women, the false positive rate of self-reported fractures was

11%, and it varied by fractured sites and education level (284). The issue of over reporting holds true for other outcomes, and the rate of false positives may vary by outcomes.

Moreover, limitations of the phenotyping method I used may also be caused by the lack of access to other data, such as clinical text (e.g., medical notes, discharge summary, etc.). If clinical text were to be incorporated in case-control definitions (e.g., by natural language processing algorithm), classification accuracy would be improved (216). In a systematic review of 19 studies, the median of sensitivity for case definition increased to 78.1% by ICD coding plus text phenotyping algorithms (including keyword searches, machine learning algorithm and rule-based algorithm) in comparison with ICD coding alone definition, whose median sensitivity was 61.7% (285). In addition, Li and colleagues identified a total of 2609 cases, while a natural language procession system found 1253 cases which were not captured by the ICD coding system (286); Baus and colleagues found that combined use of ICD-9-CM codes with text significantly increases the total case number compared with ICD-9-CM coding alone (mean = 1256.1 for ICD-9-CM plus text vs mean = 1174 for ICD-9-CM alone) for case definition of essential hypertension (287). Therefore, fully use of structured and unstructured data (including billing codes, clinical notes, medications and lab and test results) in EMR in phenotyping algorithm is recommended by previous study (288).

Lack of individual vitamin D data

A valid instrumental variable must be strongly associated with the exposure of interest, which is serum 25(OH)D concentration in my study. Although the SNPs I selected came from the largest GWAS of white ancestry, which should be associated with 25(OH)D level in the UK Biobank population as well, it would have been desirable to test this MR assumption in the UK Biobank sample. UK Biobank has measured several serum biochemical markers, including 25(OH)D. However, they have not yet released the data. If these data were available, in addition to testing the association between IV and exposure, I could also have studied the observational associations between 25(OH)D level and outcomes by PheWAS. In addition, with individual 25(OH)D level

data, a GWAS for 25(OH)D in the UK Biobank can also be considered. Since its sample size is much larger than the recently published GWAS of vitamin D, novel vitamin D related loci could possibly be identified. Once the biomarker data is released by the UK Biobank, all the above points can be considered by future studies.

Generalizability

The previous GWAS on vitamin D included participants of white ancestry but from a number of different regions across the world. However, all UK Biobank participants reside in the UK, a high latitude country with low sun exposure on average especially during winter. Due to the weather conditions in the UK and the large impact of sun exposure on vitamin D synthesis, UK population could be of possibly lower 25(OH)D than populations in lower latitude areas, especially during the winter. However, as mentioned above, serum biomarker data has not been released, so the magnitude of this difference remains unknown. As indicated by the difference of R^2 value between SOCCS individuals and previous GWAS samples (1.61% vs 2.84%), the proportion of 25(OH)D variance explained by genetic variants in UK population might not be the same as that in the original GWAS. Hence the effect size for each variant might also diverge. Therefore, the accuracy of the application of genetic variant effects sizes from the SUNLIGHT GWAS to UK Biobank population is not certain, but it cannot be explored without the individual level 25(OH)D data.

Furthermore, whether the findings from the UK population can be generalized to the broader white population is also a question. Previous studies suggested that the relationship between vitamin D concentration and health outcomes may be nonlinear. Reid and colleagues observed an interaction between 25(OH)D and treatment effect from a 2-year trial data of 452 participants. For participants with baseline 25(OH)D level of no more than 30 nmol/L, body mass mineral change was observed between treatment group and placebo group at spine and femoral sites. But when they analysed all participant irrespective of baseline 25(OH)D level, body mass mineral changes at spine and femoral sites were not statistically significant (289). Similarly, Macdonald and colleagues analysed other trial data of 305 postmenopausal women, and found that vitamin D supplementation only increased bone mineral density at spine and hip

among individuals with serum 25(OH)D of no more than 30 nmol/L level (290). Thus, the effect of 25(OH)D on health outcomes may differ by baseline serum 25(OH)D level. Considering the limited diversity of 25(OH)D levels of the UK population, my study might lack power in detecting true effect of 25(OH)D at certain levels. In addition, the limited diversity of 25(OH)D levels in the UK Biobank might cause a lower R^2 level as shown above from the SOCCS sample (i.e., a lower level of the variance in vitamin D level is explained by genetic factors), which further decrease the statistical power of analysis dramatically.

Finally, the recruitment policy of UK Biobank might potentially bring several problems. Since UK Biobank participants were invited by mail, there may be selection bias in the UK Biobank, which makes it an unrepresentative sample of the general UK population. In fact, the UK Biobank mailed 9.2 million invitations, and only 503,325 participants responded (response rate, 5.47 %) (291). Fry and colleagues compared the sociodemographic, physical, lifestyle and health-related characteristics of UK Biobank participants with those from UK national surveys. They found that UK Biobank participants were more likely to be older, to be female, and to live in less socioeconomically deprived areas compared with UK general population. In addition, UK Biobank participants were less likely to be obese, to smoke or to drink and were of fewer self-reported health conditions (292). Their study suggested that the UK Biobank participants differed from the general UK population with regards of their social economic status, lifestyle, and disease prevalence. This selection bias could potentially cause invalid results (293).

Selection bias is especially true for observational studies. In a MR design, values of instrumental variables are not expected to be associated with socio-economic status or lifestyle factors. My analysis for the association between the genetic score of the 6 SNPs or individual SNPS and age/ BMI/ time spend outdoors/ income/ qualification/ alcohol intake frequency confirmed this.

In addition, participants can withdraw the cohort without any reason at any time. When I first downloaded the UK Biobank data, data from 502,655 participants were available.

However, in the final genetic data release, only 487,411 were still available. Among them, 39 individuals withdrew their participation. The others were caused by low quality of genotyping data. As the follow up goes on, there would possibly be more in the future. It would be crucial to check the characteristics of withdrawn individuals in order to study whether drop off happens at random. Therefore, when results from the UK Biobank cohort are to be generalized to the general population of the UK, the above discussed points should be taken into consideration.

6.2.2 Strengths and limitations of the PheWAS methodology

6.2.2.1 Strengths of the PheWAS methodology

Taking advantage of EMR record, PheWAS tests associations between selected genetic variants or exposure and a wide range of phenotypes, diagnoses, traits or outcomes. In contrast with GWAS, which tests association between genetic variants and a single or a small selection of outcomes (e.g., blood pressure, BMI), PheWAS focuses on the broad associations between predetermined exposure and a comprehensive range of outcomes. As a reverse GWAS, which is hypothesis generating as well, PheWAS is an important tool for identifying new biomarkers, elucidating genetic architecture of complex traits, and uncovering pleiotropy (294). Previous PheWAS studies have successfully validated known associations, and identified novel associations as well (section 2.2.1). In addition, Moore and colleagues implemented a PheWAS for associations with 5,954,294 polymorphisms and 27 laboratory phenotypes in 2,547 individuals from human immunodeficiency virus clinical trial data. They found 10,963 nominal genotype-phenotype associations ($P < 0.01$), and 29 possibly novel association (e.g., rs10494326 with neutrophil counts and rs2201841 with plasma chloride concentration) (236). A limitation of their study might be that although they tested ~1.6 million associations, they defined their statistical significance threshold as $P < 0.01$ in both datasets and consistent direction between the two datasets (they split 4 trials into two comparable datasets to seek internal replication), rather than implementing any statistical process controlling for multiple testing burden. However, their study did suggest the possibility of PheWAS for identifying pharmacogenomic associations. Moreover, when applied in trial data, PheWAS can also be used to identify side/adverse effects of treatment factor (e.g., medication) effectively, without

any prior assumptions.

In this study, I comprehensively studied the association between 6 vitamin D related SNPs, a score of 6 SNPs and 920 outcomes. In fact, I used it in combination with Mendelian Randomisation, forming a MR-PheWAS design as the study published by Millard and colleagues (235). Compared with traditional single outcome MR studies, this design can simultaneously explore causal associations between selected predictor and great number of phenotypes.

Previous studies were mainly for establishing methods and workflow for PheWAS of different types. But they have successfully demonstrated the value of PheWAS for genomic research, including replicating known associations and identifying suggestive novel associations. It is possible that PheWAS will become a powerful tool used by both clinicians and basics scientists. Studies conducted in trial data were good examples for this (236, 243). In addition, PheWAS is also useful for clinicians in real-word environment. Since EMRs have large numbers of individuals, who were exposed to multiple drug across their lifetime, important relationships between genetic variation and pharmacological responses as well as the factor of time and age could be highlighted (295). Moreover, as the cost of developing new drugs increasing over the past 60 years, PheWAS provides attractive opportunities for drug repurposing through its high-throughput computational analysis (296). For basic physiologist, PheWAS will complement experimental research and help validate insight gained from studying genes and gene variants in model organisms (297).

In the near future, the construction of large Biobanks linked with EMRs, which also collect broad range of epidemiological phenotypes (e.g., the UK Biobank), could help generate promising findings from PheWAS. Despite its good potential, PheWAS is still in its early age. It has several limitations as listed below, which should be conquered by methodological improvements.

6.2.2.2 Limitations of the PheWAS Methodology

PheWAS heavily relies on the phenotyping method/algorithm it applied. In my study,

I employed a phecode system developed by Denny and colleagues (207). This algorithm defines cases and controls with ICD codes, which included hospital inpatient ICD codes, cancer and death registry codes. As it has been discussed in section 6.2.1.2, this phenotyping method might cause substantial issues with misclassification. It should be acknowledged that the algorithm for phenotyping is yet to be fully developed. To make the algorithm perform better, the following efforts could be considered. Firstly, incorporate self-reported data in phenotype definition. In the subsequent MR analyses, I merged EMR data with self-reported data for a selection of outcomes. The UK Biobank group is also conducting similar work as well. They adjudicate phenotypes for individuals based on their hospital in-patient data, cancer/death registry and self-reported data by algorithm. Cases were reported as prevalent cases or incident cases. Prevalent cases were reported as identified by EMR (with or without self-report) or identified by self-reported only. Incident cases were reported as identified by EMR or identified by register data. They have released adjudicated outcomes for myocardial infarction and stroke outcomes, and outcomes on more common diseases will be released in the future. In addition, leverage clinical text in case/control definition. Although large amounts of medical notes were generated, they were not fully exploited in PheWAS. With the development of deep neural networks, clinical text processing programmes are expected which manipulate text and can assign case/control status for health outcomes effectively in a high throughput way. Moreover, a good classifier could take data from multiple sources (e.g., EMR codes, clinical texts, lab data, imaging data), integrate them and give reliable judgements on the case/control status for diseases. There have been several medical information extraction algorithms based on natural language processing, for instance cTAKES and MedLEE (298). In addition, Denny and colleagues developed 13 algorithms which incorporate EMR codes with clinical text, medication and lab data (for phenotypes cataract, dementia, type 2 diabetes, diabetic retinopathy, resistant hypertension, peripheral arterial disease, primary hypothyroidism, low levels of high-density lipoprotein cholesterol and baseline lipid values, red blood cell indices, white blood cell indices, normal cardiac conduction and height) (202). However, the information extraction algorithms are criticized as being unable to capture subtle relationships hidden in clinical notes, since the language is complex and algorithms lack explicit semantic resources describing

the relationships between clinical concepts (298). The algorithms developed by Denny and colleagues focused on selected diseases, which were determined on clinical priority and their research interest. Therefore, extra efforts are still needed in algorithm development, especially in making a complex algorithm, which integrates multiple source data and defines case/control status for all diseases comprehensively.

As mentioned above, the statistical power of genotype-phenotype association depends on the number of cases for each health outcome. Although a total of 1853 phenotypes were defined through EMR data by the phecode system, only 920 outcomes had more than 200 cases and were included in the analysis. This threshold of 200 cases was based on a previous simulation for PheWAS power (299). The reasons for employing a case number threshold was to exclude low power phenotypes, release multiple testing burden, and meanwhile, to comprehensively explore associations for as many outcomes as possible.

Since large number of phenotypes are tested simultaneously in a PheWAS, there is a great multiple testing burden. In GWAS, a total of 1 million independent variants across the whole genome is examined, so applying a Bonferroni correction, a *P* value of smaller than $5*10^{-8}$ is considered as statistically significant. To guarantee the repeatability of the results, replication in independent samples is usually conducted. However in a PheWAS, phenotypes are usually correlated. Applying Bonferroni correction, which divide 0.05 by the number of tested phenotypes, is over conservative. In my study, the phecode system first binned ICD codes according to their clinical relationship, and then I applied Bonferroni correction. However, this did not guarantee that all the phecode groups were independent, so the application of Bonferroni correction in my study also had the problem of being over conservative. Alternatively, false discovery rate (FDR) could be considered in controlling for multiple testing burden. In a conventional hypothesis testing, the *P* threshold of 0.05 controls the false positive rate of being lower than 5%, which means that if one implemented 100 independent tests, there would be at most 5 false positives. FDR is the expected fraction of tests declared statistically significant in which the null hypothesis is true (300). The Benjamini and Hochberg procedure is a common choice in FDR correction.

It first ranks all *P* values in ascending order, then it finds the largest *P* value where $P_i < d * i/n$ (i denotes the rank of *P* value, d denotes the significance level (e.g., 0.05), n denotes the total number of tests). All associations with *P* values of less than this *P* value (including this value) are declared as statistically significant (301). FDR retains more statistical power compared to Bonferroni correction, but it does not take the correlation of phenotypes into consideration. Recently, a new Bayesian algorithm, which implements a Markov process, has been developed by Cortes and colleagues. They have shown that their Bayesian framework could increase statistical power by 20% and identified novel associations in the UK Biobank cohort (302). The strength of their approach is that it analyses the association between exposure and all ICD codes of every coding level (i.e., I21 (acute myocardial infarction) and I21.0 (acute transmural myocardial infarction of anterior wall) are analysed as independent codes) while the relationships between codes can be handled by the Markov process. On the other hand, this Bayesian framework is more computationally intensive compared with traditional general linear models.

GWAS has proved its good replicability, however, the replicability for PheWAS may be lower. As has been discussed above, the current algorithm for high-throughput phenotyping is imperfect (problem of misclassification). So case-control status of different populations collected by different studies/clinical centres might be biased in different directions, causing associations observed in one population not replicated in another one. There are not many previous PheWAS which use different populations or split one cohort into two parts for discovery and replication in one study. Three previous studies analysed genotype-phenotype associations in different populations, and defined associations which were consistent and significant in more than one populations as statistically significant (233, 234, 236). However, these studies only reported the associations which met their criteria of cross-population statistical significance. It could be not estimated from their study that how many associations were significant only in one population. In 2015, Bolland and colleague conducted a PheWAS in a population of 1,749,400 from the Columbia University Medical Centre. They studied association between birth month and 1688 diseases and found that 55 diseases were significantly associated with birth month (212). In their study, nine

cardiovascular related outcomes were significant, including atrial fibrillation, essential hypertension, congestive cardiac failure, angina, coronary arteriosclerosis, cardiomyopathy, pre-infarction syndrome, chronic myocardial ischemia and mitral valve disorder. In 2016, Li and colleague tried to replicate their findings from an independent population of 1,169,599 individuals from the Mount Sinai Hospital (217). Among the nine significant cardiovascular outcomes in the previous study by Bolland and colleagues, four outcomes (coronary arteriosclerosis, essential hypertension, angina, and pre-infarction syndrome) were significant and three outcomes (atrial fibrillation, cardiomyopathy, and chronic myocardial ischemia) were of consistent pattern but not significant. Association for congestive cardiac failure and mitral valve disorder were not replicated by the latter study. As the convention for GWAS, independent replication is vital for study findings drawn from EMRs (217). In addition, future developments in better phenotyping algorithms as discussed above could help in implementing PheWAS of better replicability. In my thesis, I did not have another cohort of comparable size to the UK Biobank. However, splitting the UK Biobank into different sub-populations according to the assessment centre (or genotyping array) and seeking replication is a potential choice.

6.2.3 Strengths and Limitations of Mendelian Randomization Method

6.2.3.1 Strengths of Mendelian Randomization Method

MR study is a cost-effective analogy to trials by using genetic variants as instrumental variables in testing causal associations between exposures and outcomes. In observational studies, statistically significant association could possibly exist between independent exposure and outcome variables if a confounding factor is associated with both exposure and outcome. Thus, associations identified from observational studies should not be interpreted as causal. However, genotypes of individuals are not affected by most confounding factors (e.g., social economic status, environmental exposure), since they are determined randomly at conception. By using genetic variants as instrumental variables, Mendelian Randomization is able to explore the putative causal association between exposure (e.g., 25(OH)D) and health outcomes.

Compared to RCTs, MR studies are easier to implement in large samples. The process

of randomization for a MR study happens at conception for every participant when alleles are randomly allocated into gametes, rather than as a step in the experimental design of the RCT. MR study is more cost-effective by omitting the randomization and following up steps which are essential in a trial. It can recruit population based samples and study the causal association between life-long exposure of specific factor and health outcome retrospectively. If used in large biobanks with biorepository (e.g., UK Biobank cohort), it is feasible to be implemented in very large population.

6.2.3.2 Limitations of Mendelian Randomization Method

Although MR is generally not influenced by the common confounding factors and not seriously affected by reverse causality, it does rely on several assumptions that can be hard to identify and control and has several crucial issues that need to be noted when interpreting their results. I will discuss these issues in this section.

MR assumptions

The three main assumptions that underpin the MR method are: 1) the instrumental variable is associated with exposure; 2) the instrumental variable is not associated with any confounder of the exposure-outcome association; 3) the instrumental variable is conditionally independent of the outcome given the exposure and confounders (267).

Assumption 1:

Genetic variants should be statistically significantly associated with the exposure variable to be a valid instrumental variable. Furthermore, if the variance of exposure variable explained by the instrumental variable is small, it could reduce the statistical power of MR (265). Thus, a weak instrument is an issue that should be explored by MR studies. This can be measured by the *F* statistic. *F* statistic depends on sample size, number of instruments, and R^2 (proportion of variance of exposure explained by instruments). The bias induced by weak instrument is shown to be 100/*F* percent of the observational association between exposure and outcome. Thus, instruments with

F statistic of larger than 10 (i.e., bias is around 10% of the observational estimate) are generally considered to be robust (269). However, as criticized by Burgess and colleagues, estimating bias simply on *F* statistics measured from data is not enough (303). In their study, they spilt a cohort into 16 equally sized substudies, implemented two-stage MR for each of them, and then derived the summarized causal effect estimate by fixed-effect meta-analysis. As shown by their data, *F* statistics of the 16 substudies varied from 3.4 to 22.6. Substudies with larger *F* statistics were of higher causal effect estimate values and tighter CIs, while the causal effect estimate was near 0 when the whole cohort is analysed in MR. Although strong instruments are desirable, any guidance that relies on providing a threshold, such as excluding instruments or studies if $F < 10$, may introduce more bias than it prevents (303). Therefore, issues of weak instruments should ideally be pre-specified before data collection, by specifying sample sizes, instruments and genetic models based on the best prior evidence available. In addition, adjusting for covariates in models also helps reducing weak instrument bias (303).

This issue is especially important for vitamin D, since it is highly affected by environmental factors, such as sun exposure, when compared to other traits (e.g., height, body mass index). From previous twin studies, the hereditability of 25(OH)D concentration was estimated to be ~70% (304, 305). However, the largest GWAS conducted by SUNLIGHT consortium identified 6 GWAS significant SNPs, and they explained only 2.84% of the variance of 25(OH)D. I judged the adequacy of my instruments based on the fact that the SNPs were significantly association with 25(OH)D level from GWAS and that the *F* statistic for the score was 45.96 (calculated from SOCCS sample). Since *F* statistics is correlated with sample size, the *F* statistics calculated from UK Biobank would be larger. In addition, the wide application of the 4 loci (*DHCR7*, *CYP2R1*, *GC*, *CYP24A1*) as IVs in previous MR on vitamin D and the strong associations of all 6 loci with 25(OH)D level in GWAS provide prior evidence supporting the validity of the IVs in my study. Despite this, it should be acknowledged that the very low percentage of variance explained is a major limitation of my study.

Assumption 2:

This assumption would be violated if subgroups in the study population have both different genotype frequencies and different distributions of the outcome (population stratification), or if there is an association between the genetic instruments and confounders. In my study, the genotype distributions of 3 SNPs were statistically significantly different across UK Biobank assessment centres, which is an indication of population stratification. The polymorphism rs12785878 showed a clear pattern of genotype changing with latitude, while the other two (rs10741657 and rs10745742) did not show a clear pattern. I included UK Biobank assessment centre, longitude and latitude of home address and the first 5 ancestral principle components as covariates in my MR analyses to adjust for the potential bias induced by population stratification. The cause of associations between rs10741657, rs10745742 and UK Biobank assessment centers should be further explored. This could possibly be caused by genotyping error or biased sampling, but I cannot explore this with my current data. This issue can be checked with another independent sample collected from the whole UK, or just the affected geographical areas, whose minor allele frequencies are different from other areas.

Assumption 3:

This assumption is likely to be violated when the genotype has multiple (pleiotropic) effects, and there is another pathway other than the targeted exposure (i.e., vitamin D) through which instrument can affect outcomes. PheWAS is a good way to explore pleiotropy since associations between score or SNP and all disease outcomes were tested. From my PheWAS results, only associations between rs107216707 and kidney outcomes or alveolar and parietoalveolar pneumonopathy survived Bonferroni correction, which suggested possible pleiotropic effect of rs107216707. My results did not suggest pleiotropic effect of the other 5 SNPs or the score.

When pleiotropy is present significant associations found from MR between targeted exposures and outcomes might be false positives, which are actually caused by the

causal association between another exposure and outcome. Sensitivity analysis can be considered to explore and control the existence of pleiotropy. This can be performed, as in my study, by dropping one variant in turn and implementing MR with the other variants. If the observed association in MR with all variants were dominant by any single variant of pleiotropic effect, the result excluding that variant would differ. In addition, Egger's regression could be considered in testing the existence of pleiotropy and get an unbiased effect estimate, which will be discussed below.

In order to examine the assumptions and to obtain robust causal estimates, several MR statistical methods have been proposed, including two-stage MR, MR IVW and MR Egger. I will discuss these MR methods in the following sections.

Two Stage Mendelian Randomization

The score of 6 SNPs which I constructed was a sum score, weighted by the SNPs effect estimates from the largest-published GWAS (158). In a two-stage MR, the first stage is to fit a model for the association between the IV and the exposure, which had already been done by the SUNLIGHT GWAS (158). Afterwards, the second stage is to fit the level of exposure predicted by the IV for each participant (i.e., the process of calculating the score of 6 SNPs in my study), and run a logistic regression between the genetic determined exposure level and health outcomes. Therefore, in this way, my PheWAS for the score of 6 SNPs equals the second stage in a two-stage MR. The causal estimate is the second-stage regression coefficient that is explained as the change in the outcome caused by a unit change in the exposure and the estimator is expressed as a causal relative risk or odds ratio (306, 307). However, a drawback of conducting a two-stage MR in this way is that the standard error in the first stage (association between genetic variants and exposure) is not considered, and the estimated standard error of coefficients and *P* values would be inflated. Despite this, the two-stage estimator with a logistic regression second-stage model still provides a valid test for the null hypothesis.

Inverse Variance Weighted Mendelian Randomization

Inverse variance weighted (IVW) Mendelian Randomization is a combination of the ratio method and meta-analysis. The causal estimate for a single instrument could be calculated by the ratio method, which is the coefficient of the regression of outcome on the variant divided by the coefficient of the regression of exposure on the variant. If multiple instruments are employed, the ratio estimate from each instrument is then pooled with an inverse-variance weighted meta-analysis. In an IVW MR, the IV-specific causal estimates are equivalent to the study-specific estimates in meta-analysis, and the weights are the inverse-variance weights (267). The causal effect estimate is derived by a weighted linear regression, where the residual standard error is set to one and the intercept is constrained to zero. This equals to a fixed-effect meta-analysis. It is assumed that there is no heterogeneity between individual effect estimates for each SNP when applying a fixed-effect model. When substantial heterogeneity presents, IVW MR is not recommended. In addition, IVW also relies on the three MR assumptions as listed above. If all three are met, the causal effect estimate from IVW MR is robust and unbiased. However, assumptions 2 and 3 might be violated when using multiple genetic variants as instruments. To check these assumptions, I thus applied the MR Egger method to assess whether the genetic variants have any pleiotropic effects (directional pleiotropy).

Egger's Mendelian Randomization

The Egger's MR method is developed to provide robustness against misspecification of the MR assumption 3 (268). Egger's MR assumes that the correlation between the genetic associations with the exposure and the direct effect of the genetic variants on the outcome is zero, termed the InSIDE assumption, rather than using the original assumption 3 (268). In Egger's MR, the intercept term is no longer restricted to zero as in IVW MR. If the intercept is zero (referred as 'balanced pleiotropy'), the InSIDE assumption is satisfied, and the estimates from Egger's MR should be similar to estimates from IVW MR. If the intercept differs from zero (rejection of the null hypothesis), the InSIDE assumption is violated (referred as 'directional pleiotropy'). In this situation, the IWW MR estimates are biased, but Egger's MR provides effect estimates closer to true effect compared with IVW MR (268). Hence, testing the intercept from the MR Egger analysis provides an assessment of the validity of the IV

assumption. Although the Egger's MR is more robust in dealing with pleiotropy, this method also has some limitations, which include the precision of the estimate, the influence of outlying variants, and the limited statistical power. These limitations should be noted when interpreting results from Egger's MR.

Accurate effect estimate from Egger's MR requires the consistency of the effects across the genetic variants. When pleiotropic variants were employed as IVs, heterogeneity between the causal estimate of individual variant would be observed. This would result in over-dispersion in the MR Egger regression, in which a random-effects model is preferred. Therefore, the standard error of the causal estimate from the MR Egger method (random-effects model) is typically larger than that from the MR IVW method (fixed-effect model) and accordingly the 95%CI of the causal estimate from the MR Egger method is also wider than that from the MR IVW, which would result in an imprecise estimate (308).

In addition, the MR Egger estimate is easily influenced by an outlying variant (268). If one variant has a much stronger association with the exposure than others, this variant would have a larger influence on the coefficients in the Egger's MR compared with other variants. Thus, the causal effect estimate from Egger's MR would be dominated by this outlying variant. Among the 6 variants I employed, rs3755967 (GC) had a much larger effect on 25(OH)D concentration than other variants (shown in figure 37). This implied that the Egger's MR results might be biased by this variant. Hence, I implemented sensitivity analyses afterwards in which I excluded rs3755967 and only used the other 5 variants. But the results were not changed by excluding rs3755967.

Finally, Egger's MR is of limited power, which is a common issue shared by all MR methods. The statistical power of MR methods depends on variance of exposure explained by the IVs, estimated effect size, number of cases, case/control ratio and the *P*-value threshold. The number of cases impacts the statistical power a lot. In my study,

I analysed 920 outcomes in the PheWAS using a score of 6 SNPs. This was based on a rough power estimation for PheWAS, rather than a power estimation for an MR. If I assumed a R^2 of 0.03 and a control/case ratio of 5 or larger, phenotypes with more than 9000 cases would have a power of more than 80% for detecting an OR of 1.2 or larger at a 0.05 alpha level. Based on this calculation, only 9 outcomes had sufficient power in MR analysis, including systolic blood pressure, diastolic blood pressure, risk of hypertension, risk of type 2 diabetes, risk of ischaemic heart disease, body mass index, risk of depression, risk of non-vertebral fracture, and all-cause mortality. Although the PheWAS for the score equalled a two-stage PheWAS, the high *P* values for outcomes with small sample size did not necessarily mean that there is not causal association between 25(OH)D concentration and the outcome due to the limited statistical power for low-sample-size phenotypes.

6.3 Findings from the thesis

In this part of the chapter, the following results from my thesis will be presented and discussed: 1) Genotypic distribution; 2) Confounding factors; 3) PheWAS results; 4) MR results.

6.3.1 Genotypic distribution

I first checked whether genotype frequencies of the selected variants distribute evenly across different assessment centres. A significant difference might imply population stratification, which violate the MR assumption. The frequencies of rs3755967, rs8017720 and rs17216707 were distributed evenly across all UK Biobank recruitment centres. However, for the other three SNPs, including rs10741657, rs12785878 and rs10745742, genotypes from UK Biobank assessment centre differ significantly. For the rs12785878, participants recruited from Edinburgh and Glasgow had a G allele frequency of 0.191 and 0.190, respectively. However, participants from middle England had a slightly higher allele frequency. The frequencies for G allele from Leeds and Manchester were 0.207 and 0.204, respectively. Centres from south UK had even higher frequencies. For instance, the G allele frequencies for participants from Oxford and Bristol were 0.223 and 0.218, respectively. When I pooled participants from Scotland together (i.e. Edinburgh centre and Glasgow centre), and tested the genotype

counts of Scotland versus all other centres, the *P* value for difference was lower than $2.2*10^{-16}$. This showed an uneven distribution of genotype across latitudes, which was consistent with data reported by a previous study (309). In this study with 6,877 participants from the 1958 British birth cohort, authors tested the associations between five vitamin D related SNPs (rs4588 (*GC*), rs12785878 (*DHCR7/NADSYN1*), rs10741657 (*CYP2R1*), rs6013897 (*CYP24A1*), rs10877012 (*CYP27B1*)) and multiple confounding factors, including region, sun cover, time spent outside, physical activity, oily fish intake, socioeconomic class at age of 42 years (309). In their study, the geographical region of residence was based on Government Office Regions, and were classified into South (South East, South West, and Greater London), Middle (East Anglia, Midlands, and Wales), North (North, North West, and Yorkshire and the Humber), and Scotland. When used as an outcome in the analysis, region was dichotomized into South/Middle vs North/Scotland. From their analyses, rs12785878 (*DHCR7/NADSYN1*) was statistically significantly associated with geographical region ($P = 3.0*10^{-5}$). This genotype difference is an indication of population stratification. This may be caused by various reasons, including drift and selection. Drift and selection are evolutionary processes which would affect genetic variation within and among populations (310). The actual cause and implication of this allele frequency difference across latitude needs to be explored by further population genetics studies.

Although rs10741657 was associated with assessment centre, it did not show any association with region or location. It seemed that only participants from Wrexham (Wales, n = 490, 0.562) and Cardiff (Wales, n = 12,781, 0.587) were of lower G allele frequencies compared with participants from other assessment centres (*P* value for chi square test was 0.001). When I pooled participants from Edinburgh and Glasgow together, the *P* value for participants from Scotland vs England/Wales was 0.878, suggesting that the distribution of rs10741657 genotype does not differ by latitude. This is consistent with the results from the 1958 British birth cohort study. In their study, rs10741657 was not associated with region either (311). From another MR study of Scottish sample, the G allele frequency for rs10741657 was 0.608, which was similar to the allele frequency of Scottish participants in this study (190). Alternatively,

this difference was not well explained by population stratification (Wales population vs all others) or genotype distribution across longitude. The third assessment centre in Wales was Swansea, whose allele G frequency was 0.604 (n = 1,620), even higher than those centres in England and Scotland. Meanwhile, Swansea is the most western centre, which is of higher allele frequency, while Wrexham and Cardiff locate between Swansea and all other centres but were of relative low allele frequencies. The genotype distribution of this polymorphism needs to be further explored by other studies of independent samples.

In addition, rs10745742 (*AMDHD1*) was also found to be associated with assessment centre in this analysis. The distribution of its genotype did not show any gradient across geographical region either. Participants from Edinburgh had similar genotype frequency to those from southern centres (C allele frequency 0.628), which was a large difference from those participants from Glasgow centre (C allele frequency 0.643). In fact, participants from Cardiff (Wales, n = 12,698, 0.639), Glasgow (Scotland, n = 12,613, 0.643), Wrexham (Wales, n = 482, 0.658) and Swansea (Wales, n = 1,611, 0.661) were of higher C allele frequency compared with participants in other centres. The *P* value for Cardiff/Glasgow/Wrexham/Swansea vs other centres were smaller than $2.2*10^{-16}$. HapMap and 1000 Genomes have reported its C allele frequencies to be 0.58 (n = 1,006, 1000 Genomes, EUR) and 0.55 (n = 104, HapMap, CEU) in European ancestry, respectively, which were close to the majority of assessment centres except for Cardiff, Glasgow, Wrexham and Swansea. I did not find any other previous studies reporting the allele frequency of this SNP in human of European ancestry. The genotype distribution of rs10745742 in the UK need to be further studied by other studies.

The polymorphism rs12785878 is the only SNP showing clear evidence of population stratification, which might invalidate the variants as a IV and bias the results. In my PheWAS and MR analysis the first 5 PCs, assessment centre and latitude/longitude of home address were included in the analysis. Thus, the effect of population stratification should be controlled. Furthermore, I also conducted sensitive analysis excluding rs12785878. Relevant results will be further discussed in section 6.3.3.

6.3.2 Confounding factors

In addition to UK Biobank assessment centre, I also investigated whether the score or individual SNPs were associated with other common confounding factors, including age, BMI, time spend outdoors in summer, time spend outdoors in winter, average household income, educational qualification and alcohol intake frequency.

The score was not associated with any confounding factors except for assessment centre, implying that the score was a good instrumental variable for MR analyses. Since first 5 PCs, assessment centre, and home address was adjusted in my analysis models, the PheWAS and MR results for the score should be free from these common confounding factors.

The variant rs10741657 (*CYP2R1*) was associated with BMI ($P = 0.038$) and sex ($P = 0.036$) in general. This shows that the genotype frequency diverges in different genders. However, since the P value is close to 0.05, this association is only nominal significant. The difference in genotype frequency might be brought about by sampling bias. When stratified by sex, this variant is not associated with BMI ($P = 0.212$ in females; $P = 0.156$ in males).

The variant rs12785878 (*DHCR7*) is associated with educational qualifications in both genders and females ($P = 0.046$ in both genders; $P = 0.024$ in females). Through literature search on PubMed and search on the GWAS catalogue, I did not find previous studies reporting any association between rs12785878 and educational qualifications. It was only reported by GWAS to be associated with 25(OH)D level. The variant rs17216707 is associated with alcohol intake frequency in males ($P = 0.022$). Through a literature search, rs17216707 was reported to be associated with glomerular filtration rate and vitamin D level by previous studies. There has not been any study reporting its association with alcohol intake. Since the P values for rs12785878 with educational qualifications and for rs17216707 with alcohol intake in males were only nominal from my results, my thesis merely provided any evidence supporting these associations.

At last, the variant rs3755967 (*GC*), rs10745742 (*AMDHD1*), rs8018720 (*SEC23A*) was not associated with any common confounding factor except for assessment centre.

6.3.3 Findings from PheWAS analyses

I first implement PheWAS for each of the 6 variants individually. Single-variant phenome wide analyses might reveal health outcomes associated with each variant, and evidence of pleiotropy. In addition, it might identify associations between single variant and outcomes which is independent of the vitamin D pathway. In these single-variant PheWAS, I did not find any significant or suggestive association for rs12785878 (*DHCR7*) and rs10745742 (*AMDHD1*). Results for the other 4 SNPs will be further discussed below.

Results of the PheWAS for rs3755967 (GC)

None of the phenotypes were statistically significantly associated with rs3755967 at a *P*-value of $5.44*10^{-5}$ (i.e., after applying Bonferroni correction). It was found to be suggestively associated with pilonidal cyst (overall and in males), and with otitis externa (in females) at *P* level of smaller than 0.001. Pilonidal cyst is a type of cyst filled with hair and skin debris, which always occurs in the lower back. There is not any previous study which has reported an association between *GC* or rs3755967 and pilonidal cyst. Otitis externa is caused by inflammation of the ear canal, which often presents with ear pain, swelling of the ear canal, and sometimes decreased hearing. I did not find any previous study exploring association between *GC* or rs3755967 and otitis externa. In addition, by searching the GWAS catalogue, there have not been any outcomes reported to be associated with rs3755967 other than vitamin D level. Since my MR analysis did not support any association between 25(OH)D level and pilonidal cyst or otitis externa, the suggestive association observed by me was unlikely to be caused by vitamin D levels. This might be caused by other biological pathways linked with *GC* gene or other nearby mutation or genes which are correlated with rs3755967 in surrounding area. Future studies with sequencing data or fine mapping of surrounding area may be helpful to explore the reason of this observed associations.

Results of PheWAS for rs10741657 (CYP2R1)

In a pooled PheWAS including both sexes, rs10741657 was not statistically associated with any outcome. However, its associations with nephrotic syndrome, labyrinthitis and complications of cardiac/vascular device, implant, and graft were suggestive at a $P < 0.001$ level. Nephrotic syndrome refers to excessive proteinuria, with associated hypoalbuminemia, edema and hyperlipidemia (312). Labyrinthitis is an inner ear infection, which causes labyrinth to be inflamed and affects hearing and balance. CYP2R1 is involved in the synthesis of cholesterol, steroids and other lipids. The mechanism linking CYP2R1 to the above three outcomes is unknown, and there have not been studies reporting associations between *CYP2R1* and those three outcomes. Through a GWAS catalogue search, this variant has only been reported to be associated with vitamin D level and vitamin D deficiency by previous GWAS (158, 159).

In males, none of the tested outcomes survived Bonferroni correction, but associations with premature cardiac beats and carcinoma in situ of skin were suggestive at a $P < 0.001$ level. A premature beat is an extra heartbeat resulting from abnormal electrical activation originating in the ventricles before a normal heartbeat would occur. Similarly, there is not any previous evidence (including any published GWAS) linking CYP2R1 with premature beats or carcinoma in situ of skin in either gender. These associations suggested by my study may need to be further explored by other studies. Future cross-sectional studies with sequencing data or fine mapping of the surrounding region, or cell/animal studies may help to unravel the true biological pathways causing these associations.

Results of PheWAS for rs8018720 (SEC23A)

The polymorphism rs8018720 was not statistically significantly associated with any outcome. However, the results suggested associations with “myalgia and myositis” and “rheumatic disease of the heart valves” at $P < 0.001$ level. There have not been any previous publications reporting associations between rs8018720 or gene *SEC23A* and myalgia, myositis or rheumatic disease of the heart valves. There is not any previous GWAS studies reporting association between rs8018720 and outcomes other than vitamin D level. The encoded protein SEC23A was suggested to be involved in

ER-Golgi protein trafficking (313). The mechanisms linking *SEC23A* with the above three outcomes, if any, may need to be elucidated by future studies.

Results of PheWAS for rs17216707 (CYP24A1)

The associations between rs17216707 and "calculus of ureter", "urinary calculus", "alveolar and parieto-alveolar pneumonopathy" survived Bonferroni correction in the analysis including participants of both sexes. Associations between rs17216707 and "calculus of kidney" were suggested in both sexes and females, and associations with "urinary calculus" and "calculus of ureter" were suggested in females and males at a *P* level of less than 0.001.

As described in section 1.3, the gene *CYP24A1* encodes 25(OH)D-24-hydroxylase which hydroxylase the side chain of $1,25(OH)_2D$ and degrade it to $24,25(OH)_2D$ (165). In addition, $1,25(OH)_2D$ can be finally transformed into calcitroic acid, which is secreted in bile. Thus, it plays an important role in calcium homeostasis. There have been previous studies reporting associations between mutations in *CYP24A*1 and hypercalcemia, hypercalciuria, and kidney diseases (e.g., nephrolithiasis and nephrocalcinosis) (314, 315). Previous studies were case reports, so my study observed the association between *CYP24A1* and renal stones at population level for the first time with a very large sample size. My results suggest that the association between *CYP24A1* and renal outcomes might not be caused by the causal effect of 25(OH)D. Gene *GC* explained a larger variance of 25(OH)D concentration compared to *CYP24A1*, but the associations between *GC* variant and renal outcomes were at a *P* value of greater than 0.001 level. In addition, in the MR analysis, when I combined all 6 SNPs together, the association between vitamin D and renal outcomes was not statistically significant. In a previous study it was suggested that in chronic kidney disease patients, elevated serum phosphate and fibroblast-like growth factor 23 levels would increase CYP24A1 expression, which causes downstream vitamin D deficiency and contributes to other complications of renal disease (316). The role of *CYP24A1* in risk of kidney stones and other related outcomes, which is independent of 25(OH)D level, deserves further study by future laboratory, animal and other population level studies.

In short, by single-variant phenome wide analyses, novel associations for the 4 SNPs were suggested, which provide evidences for future studies.

6.3.4 Findings from MR analyses

Following the single-variant phenome wide analyses for each of the 6 SNPs individually, I explored causal associations between 25(OH)D level as an exposure and broad health outcomes with a PheWAS for score of 6 SNPs and further MR analyses. Since the PheWAS used a score of 6 SNPs weighted by their effect estimates from previous GWAS, which equals the 25(OH)D level predicted by genetic variants, the results from the PheWAS equalled a two-stage MR. In the PheWAS, none of the phenotypes survived Bonferroni correction. Therefore, the PheWAS implied that there was no causal effect of vitamin D on all the 920 outcomes tested.

Although with a limited cases number (n=291), vitamin D deficiency came out as the outcome with the third smallest *P* value among all the phenotypes I tested (*P*=0.00116) in the PheWAS of score, suggesting the validity of the instrument and methodology I applied. Since the EMR data I used in the PheWAS was from hospital inpatient data, cancer registry and death registry, I suspect that a large number of cases for vitamin D deficiency were missed by the EMR dataset. Because this condition is largely under-detected and not normally treated in hospital, and participants would not commonly be aware of its presence.

At last, in the final MR analyses, I merged the EMR data I used in PheWAS with self-reported medical conditions in case/control classification, in order to capture cases missed by EMR data. I merged phenotype data for nine outcomes, and conduct MR analyses of three different methods (Table 52). Results from the MR analyses will be discussed and compared with previous MR studies below.

Evidence from a previous MR study by Kunustor et al. did not find any significant association between vitamin D (instrument of 4 SNPs) and SBP or DBP in a sample of 69,395 individuals (176). Another study by Skabby et al. also did not observe

significant association (178). However, in a MR study using summary data, involving up to 50 individual studies, DBP and risk of hypertension were found to be statistically significantly associated with vitamin D in a sample of more than 140k participants (DBP: -0.29 mmHg, 95% CI: -0.52 to -0.07, $P = 0.01$; risk of hypertension: OR = 0.92, 95% CI: 0.87 to 0.97, $P = 0.002$) (179). In my study, by merging the self-reported data, I had SBP and DBP values for 319,778 participants, and 106,405 cases of hypertension. The estimated power for detecting an effect with an odds ratio of 1.2, assuming explained variance of 0.03 of 25(OH)D level, is 1.00. However, the effect estimates from my analysis were -0.648 mmHg (s.e.=0.451, P=0.210) for SBP, -0.117 mmHg (s.e.=0.251, P=0.661) for DBP, and 0.973 (95% CI: 0.911-1.040, P=0.340) for risk of hypertension per standard deviation increase of log transformed vitamin D level by IVW MR. Since I used participants from the same study, the phenotype collections (blood pressure measurement, and hypertension diagnosis and coding) were measured consistently across the whole cohort. As noted above due to the study design characteristics of UK Biobank data from this study might be expected to be less vulnerable to the problem of population stratification, which may be an issue for MR based on summary statistics and may cause false positives. In accordance to two previous MR studies (176, 178), my study suggests that there is no moderate to large causal effect (greater than 1.2) of vitamin D on blood pressure outcomes and risk of hypertension.

Type 2 diabetes has a high global prevalence and the association between vitamin D and T2D was summarized as suggestive by evidence from previous observational studies and RCTs (25). Most of the previous MR studies did not support any causal effect of vitamin D on risk of T2D (170, 173, 317), except for one (171). In my study, I had 15,958 cases for T2D and a 97% estimated power of detecting an effect of 1.2. However, the observed OR estimation for risk of T2D was 0.971 (95% CI: 0.845-1.117, P=0.971) per standard deviation increase of log transformed vitamin D level. Comparing results from the present study and results from previous MR studies, I provided evidence arguing a null causal effect of vitamin D and T2D.

Similarly, the association between vitamin D and cardiovascular diseases was also

suggestive by the previous umbrella review (25). However, none of the previous MR studies observed any causal effect (175, 177). In my phenotype data incorporating EMR and self-reported data, I had 28,337 cases for ischemic heart disease and 1.00 power of detecting an OR of 1.2. The estimated OR from my study was 1.020 (95% CI: 0.971-1.135, *P*=0.647) per standard deviation increase of log transformed vitamin D level for IHD. Thus, my study did not support any causal effect of vitamin D on IHD.

From the previous umbrella review summarizing evidence from observational studies and RCTs, the association between 25(OH)D level and BMI was suggestive (25). However, two previous MR studies did not find any statistically significant association (173, 195). In my sample of UK Biobank cohort, the estimated effect was 0.130 (SD=0.121, *P*=0.329) kg/m^2 per SD increase of log transformed vitamin D level. Thus, current evidence from MR analyses does not support a causal effect of vitamin D on BMI.

From the previous umbrella review summarizing evidence from observational studies and RCTs, the association between 25(OH)D level and risk of depression was suggestive (25). However, to the best of my knowledge this association has never been tested in a MR study. My study is the first one to explore the effect of vitamin D on risk of depression in a large prospective cohort. In the merged dataset of EMR data and self-reported data, I had 23,294 cases for depression, and 1.00 power of detecting a true effect of 1.2. However, the estimated causal OR from IVW MR was 0.913 (95% CI: 0.816-1.022, *P*=0.093). My study does not support any causal association between vitamin D and risk of depression.

The first biological function vitamin D was found to be involved in was calcium absorption and metabolism (318) and it is widely believed vitamin D would be causally associated with bone related outcomes. According to the previous umbrella review, the association between vitamin D and non-vertebral fractures was concluded as suggestive (25). From my systematic literature review of previous MR studies on vitamin D, I did not find any MR study on non-vertebral fracture. However, I found

one on bone mineral density (BMD) and bone metabolism biomarkers. Li et al. explored associations between 25(OH)D level and BMD at the lumbar spine, BMD at femoral neck, total hip BMD, parathyroid hormone and procollagen type 1 N-terminal pro-peptide in 1,824 postmenopausal Chinese women with four 25(OH)D related variants as IVs. In their observational analysis, individual vitamin D levels showed a statistically significant association with all the above outcomes. However, when they used the four SNPs as IVs and conducted MR analyses with two-stage least square model, none of the association retained statistical significance (193). I tested the causal association between vitamin D IVs and risk of non-vertebral fracture with 23,603 cases in UK Biobank data (power = 1.00, for an OR of 1.2). The OR estimation by IVW MR from my study was 0.969 (95% CI: 0.867-1.083, *P*=0.495). The results from my study do not support any role of vitamin D in risk of non-vertebral fractures. To investigate the effect of calcium, vitamin D or combination of calcium and vitamin D on incidence of fracture, Zhao et al. conducted a systematic literature review. They included 33 randomized trials with a total of 51,145 participants, which compared calcium, vitamin D or combined calcium and vitamin D supplements with a placebo or no treatment for fracture incidence in community-dwelling adults older than 50 years. Specific outcomes they tested included hip fracture, non-vertebral fracture, vertebral fracture and total fracture. However, they did not find statistically significant associations for most of the fracture outcomes. From their meta-analyses, the RR was 0.95 (95% CI: 0.82-1.11, *P*=0.54), 1.10 (95% CI: 1.00-1.21, *P*=0.05) and 0.88 (95% CI: 0.75-1.03, *P*=0.10) for calcium, vitamin D and calcium plus vitamin D supplementation on incidence of non-vertebral fracture, respectively (319). To summarise, results from my MR study with UK Biobank did not show any causal link between vitamin D and the risk of non-vertebral fractures, and similarly a recent systematic review summarizing the evidence from previous trials did not support effect of calcium or vitamin D on incidence of fractures.

Finally, all-cause mortality was tested in MR analyses. In a previous MR, Afzal et al. created a score of 4 SNPs as IV and studied its association with mortality outcomes in 95,766 white Danish (10,349 cases for all-cause mortality). For all-cause mortality, the OR estimation was 1.3 (1.05-1.61) per 20 nmol/L lower 25OHD level (187). In UK

Biobank cohort, I had a total of 9,830 deaths in my post QC white British sample, and an 87% power detecting a true OR of 1.2. From the IVW MR, the estimated OR was 1.030 (95% CI: 0.907-1.175, *P*=0.671). Although previous MR found a statistically significant effect of vitamin D for all-cause mortality with an OR of 1.3, I have shown that I have enough power at an OR of 1.2 with the UK Biobank sample. The inconsistency between the two studies could be further explored by future studies.

Chapter VII. Conclusions and Recommendations

In this last chapter, I will draw the main conclusions of my thesis. In addition, recommendations for future studies according to my findings will be presented and discussed.

7.1 Conclusions

The work I have done in this thesis was comprised of three parts. A systematic literature review of previous studies; a PheWAS analysis, which explores the association between genetic variants and hundreds of outcomes; and a MR analysis, which explored the causality between 25(OH)D level and selective outcomes. From systematic literature reviews (updated umbrella review for vitamin D, review for PheWAS studies and review for MR studies on vitamin D), I showed that vitamin D have been associated with numerous outcomes, but with conflicting evidence from previous studies. The PheWAS suggest possible novel associations between single variants and outcomes, and identify significant associations between *CYP24A1* and renal outcome, which is independent of the vitamin D pathway. The MR analyses did not support causal associations between vitamin D level and any health outcomes with large sample size and good statistical power.

7.1.1 Main conclusions from systematic literature reviews

The first vitamin D umbrella review was published in 2013. Since then there have been 95 new meta-analyses of observational studies or RCTs. The research area of studying the association between vitamin D and non-skeletal outcomes keeps growing fast. For the systematic literature review on MR, there have been only 29 studies, which met my inclusion criteria. Although associations between vitamin D and several health outcomes were probable or suggestive based on evidence from observational studies or RCTs, the associations were poorly replicated by MR studies. Considering the consistency of evidence from observational studies, MR studies and RCTs, convincing evidence on causal associations between vitamin D and outcomes is still lacking. In addition, only a small number of outcomes have been studied by MR. As a method to establish causal relationship, it is still not widely used in vitamin D studies.

7.1.2 Main conclusions for PheWAS analysis

Although around 900 outcomes were tested for a weighted score of 6 SNPs and individual SNPs, there were no statistically significant associations, except for the associations between rs17216707 (*CYP24A1*) and kidney stone related outcomes. My results implicate that gene *CYP24A1* may be associated with stone formation in kidney, and this association is independent of effect of 25(OH)D level.

7.1.3 Main conclusions for MR analysis

The results from MR analyses suggested that there was no evidence of large to moderate (OR>1.2) causal associations of vitamin D on a very wide range of health outcomes. I provided evidence of greater power for nine outcomes, including SBP, DBP, risk of hypertension, risk of type 2 diabetes, risk of ischemic heart disease, body mass index, risk of depression, risk of non-vertebral fracture, risk of all-cause mortality, for which I merged the EMR data and self-reported medical data. Results from different MR methods were consistent. Egger's regression shows that there is no evidence of unbalanced pleiotropy. Excluding any single SNP or using only 3 SNP which were not associated with UK Biobank assessment centre did not change the results. These all supported that the results of null finding are robust.

Furthermore, even larger studies, probably involving the joint analysis of data from several large biobanks with extra IVs that explain a higher proportion of the trait variance, will be required to exclude smaller causal effects, which could have public health importance because of the high prevalence of low vitamin D levels in some populations.

Although MR assumes a linear model in general, methods which can deal with non-linear effects are in development. Silverwood and colleagues proposed a method based on estimating local average treatment effects for discrete levels of exposure ranges, then testing for a linear trend in those effects (320). Future MR on vitamin D which exploit non-linear models are needed.

7.2 Recommendations

When the UK Biobank releases the general practice data in near future, it will be valuable to merge general practice data with inpatient data and registry data. This will help identify more cases and meanwhile allow inclusion of diseases that are rarely admitted to hospital. Since the general practice data use a different system from that of inpatient data, extra efforts will be needed to harmonise those data. In addition, an algorithm which could deal with the different coding structure of EMR data and self-reported data and merge them automatically for all outcomes would be helpful. As the UK Biobank follow up goes on, more incident cases will occur, which will increase statistical power for outcomes. With all the efforts above, more cases will be identified in the future which should result in increasing statistical power over time. Thus, another future PheWAS on vitamin D with the UK Biobank cohort may reveal novel findings or give more certainty about the lack of associations with health outcomes.

My study studied only white British individuals. To study the effect of vitamin D more comprehensively, PheWAS and MR study in cohorts of comparable size but with different populations (e.g., residing at lower latitude areas, having different lifestyles, sun exposure, diets) is needed, since vitamin D exposure is highly related to latitude, sun exposure and other factors. By comparing results from populations of distinct geographical regions, the differentiation of 25(OH)D effect and its underlying biological pathway can be understood.

Future genetic studies are also needed, since a big proportion of the heritability of 25(OH)D is still missing. Missing heritability has been an issue for GWAS of all traits. Firstly, GWAS of larger sample size is needed (e.g., GWAS of 25(OH)D in the UK Biobank cohort). GWAS of larger sample is of greater power, which can identify loci of tiny effects which could not be found by smaller GWAS. In addition, the missing heritability might also be explained by rare variants or gene-environment interaction. Hence, sequencing studies and interaction studies are also needed. When more loci associated with 25(OH)D level are discovered, stronger IV explaining larger proportion of 25(OH)D variance can be constructed. Then more powerful MR study would be feasible in order to detect small or even minimal causal effect.

References

1. Heaney RP. Vitamin D in health and disease. Clinical journal of the American Society of Nephrology : CJASN. 2008;3(5):1535-41.

2. [updated 14/05/2015. Available from: http://www.nhs.uk/Conditions/vitamins-minerals/Pages/Vitamin-D.aspx.

3. Wacker M, Holick MF. Sunlight and Vitamin D: A global perspective for health. Dermato-endocrinology. 2013;5(1):51-108.

4. Basit S. Vitamin D in health and disease: a literature review. British journal of biomedical science. 2013;70(4):161-72.

5. Holick MF. High prevalence of vitamin D inadequacy and implications for health. Mayo Clinic proceedings. 2006;81(3):353-73.

6. Matsuoka LY, Wortsman J, Haddad JG, Kolm P, Hollis BW. Racial pigmentation and the cutaneous synthesis of vitamin D. Archives of dermatology. 1991;127(4):536-8.

7. Wortsman J, Matsuoka LY, Chen TC, Lu Z, Holick MF. Decreased bioavailability of vitamin D in obesity. The American journal of clinical nutrition. 2000;72(3):690-3.

8. Koutkia P, Lu Z, Chen TC, Holick MF. Treatment of vitamin D deficiency due to Crohn's disease with tanning bed ultraviolet B radiation. Gastroenterology. 2001;121(6):1485-8.

9. Holick MF, Matsuoka LY, Wortsman J. Age, vitamin D, and solar ultraviolet. Lancet (London, England). 1989;2(8671):1104-5.

10. Cheng JB, Motola DL, Mangelsdorf DJ, Russell DW. De-orphanization of cytochrome P450 2R1: a microsomal vitamin D 25-hydroxilase. The Journal of biological chemistry. 2003;278(39):38084-93.

11. Tangpricha V, Flanagan JN, Tseng LWWCC, Chen TC, Holt PR, Lipkin MS, et al. 25-hydroxyvitamin D-1α-hydroxylase in normal and malignant colon tissue. The Lancet. 2001;357(9296):1673-4.

12. Broe KE, Chen TC, Weinberg J, Bischoff-Ferrari HA, Holick MF, Kiel DP. A higher dose of vitamin d reduces the risk of falls in nursing home residents: a randomized, multiple-dose study. Journal of the American Geriatrics Society. 2007;55(2):234-9.

13. DeLuca HF. Overview of general physiologic features and functions of vitamin D. The American journal of clinical nutrition. 2004;80(6 Suppl):1689s-96s.

14. Yamamoto M, Kawanobe Y, Takahashi H, Shimazawa E, Kimura S, Ogata E. Vitamin D deficiency and renal calcium transport in the rat. The Journal of clinical investigation. 1984;74(2):507-13.

15. Okazaki T, Igarashi T, Kronenberg HM. 5'-flanking region of the parathyroid hormone gene mediates negative regulation by 1,25-(OH)2 vitamin D3. The Journal of biological chemistry. 1988;263(5):2203-8.

16. Cao X, Ross FP, Zhang L, MacDonald PN, Chappel J, Teitelbaum SL. Cloning of the promoter for the avian integrin beta 3 subunit gene and its regulation by 1,25-dihydroxyvitamin D3. The Journal of biological chemistry. 1993;268(36):27371-80.

17. Polly P, Carlberg C, Eisman JA, Morrison NA. Identification of a vitamin D3 response element in the fibronectin gene that is bound by a vitamin D3 receptor homodimer. Journal of cellular biochemistry. 1996;60(3):322-33.

18. Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. The American journal of clinical nutrition. 2006;84(1):18-28.

19. Wahl DA, Cooper C, Ebeling PR, Eggersdorfer M, Hilger J, Hoffmann K, et al. A global representation of vitamin D status in healthy populations. Archives of osteoporosis. 2012;7:155-72.

20. Ginde AA, Liu MC, Camargo CA, Jr. Demographic differences and trends of vitamin D insufficiency in the US population, 1988-2004. Archives of internal medicine. 2009;169(6):626-32.

21. Samuel L, Borrell LN. The effect of body mass index on adequacy of serum 25-hydroxyvitamin D levels in US adults: the National Health and Nutrition Examination Survey 2001 to 2006. Annals of epidemiology. 2014;24(10):781-4.

22. Chapuy MC, Preziosi P, Maamer M, Arnaud S, Galan P, Hercberg S, et al. Prevalence of vitamin D insufficiency in an adult normal population. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 1997;7(5):439-43.

23. Karonova T, Andreeva A, Nikitina I, Belyaeva O, Mokhova E, Galkina O, et al. Prevalence of Vitamin D deficiency in the North-West region of Russia: A cross-sectional study. The Journal of steroid biochemistry and molecular biology. 2016.

24. Hypponen E, Power C. Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. The American journal of clinical nutrition. 2007;85(3):860-8.

25. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ (Clinical research ed). 2014;348:g2035.

26. Caini S, Boniol M, Tosti G, Magi S, Medri M, Stanganelli I, et al. Vitamin D and melanoma and non-melanoma skin cancer risk and prognosis: a comprehensive review and meta-analysis. Eur J Cancer. 2014;50(15):2649-58.

27. Liao Y, Huang JL, Qiu MX, Ma ZW. Impact of serum vitamin D level on risk of bladder cancer: a systemic review and meta-analysis. Tumour Biol. 2015;36(3):1567-72.

28. Zhang H, Wen X, Zhang Y, Wei X, Liu T. Vitamin D Deficiency and Increased Risk of Bladder Carcinoma: A Meta-Analysis. Cellular Physiology and Biochemistry. 2015;37(5):1686-92.

29. Li M, Chen P, Li J, Chu R, Xie D, Wang H. Review: the impacts of circulating 25-hydroxyvitamin D levels on cancer patient outcomes: a systematic review and meta-analysis. Journal of Clinical Endocrinology & Metabolism. 2014;99(7):2327-36.

30. Kim Y, Je Y. Vitamin D intake, blood 25(OH)D levels, and breast cancer risk or mortality: a meta-analysis. Br J Cancer. 2014;110(11):2772-84.

31. Mohr SB, Gorham ED, Kim J, Hofflich H, Garland CF. Meta-analysis of vitamin D sufficiency for improving survival of patients with breast cancer. Anticancer Res. 2014;34(3):1163-6.

32. Maalmi H, Ordonez-Mena JM, Schottker B, Brenner H. Serum 25-hydroxyvitamin D levels and survival in colorectal and breast cancer patients: systematic review and meta-analysis of prospective cohort studies. Eur J Cancer. 2014;50(8):1510-21.

33. Mohr SB, Gorham ED, Kim J, Hofflich H, Cuomo RE, Garland CF. Could vitamin

D sufficiency improve the survival of colorectal cancer patients? Journal of Steroid Biochemistry & Molecular Biology. 2015;148:239-44.

34. Khayatzadeh S, Feizi A, Saneei P, Esmaillzadeh A. Vitamin D intake, serum Vitamin D levels, and risk of gastric cancer: A systematic review and meta-analysis. Journal of Research in Medical Sciences. 2015;20(8):790-6.

35. Wang W, Li G, He X, Gao J, Wang R, Wang Y, et al. Serum 25-hydroxyvitamin D levels and prognosis in hematological malignancies: a systematic review and meta-analysis. Cellular Physiology & Biochemistry. 2015;35(5):1999-2005.

36. Liu J, Dong Y, Lu C, Wang Y, Peng L, Jiang M, et al. Meta-analysis of the correlation between vitamin D and lung cancer risk and outcomes. Oncotarget. 2017;8(46):81040-51.

37. Feng Q, Zhang H, Dong Z, Zhou Y, Ma J. Circulating 25-hydroxyvitamin D and lung cancer risk and survival: A dose-response meta-analysis of prospective cohort studies. Medicine. 2017;96(45):e8613.

38. Hu MJ, Zhang Q, Liang L, Wang SY, Zheng XC, Zhou MM, et al. Association between vitamin D deficiency and risk of thyroid cancer: a case-control study and a meta-analysis. Journal of Endocrinological Investigation. 2018:1-12.

39. Zhang Z, Yang Y, Ng CY, Wang D, Wang J, Li G, et al. Meta-analysis of Vitamin D Deficiency and Risk of Atrial Fibrillation. Clinical Cardiology. 2016;39(9):537-43.

40. Xu MY, Cao B, Yin J, Wang DF, Chen KL, Lu QB. Vitamin D and Graves' disease: a meta-analysis update. Nutrients. 2015;7(5):3813-27.

41. Del Pinto R, Pietropaoli D, Chandar AK, Ferri C, Cominelli F. Association Between Inflammatory Bowel Disease and Vitamin D Deficiency: A Systematic Review and Meta-analysis. Inflamm Bowel Dis. 2015;21(11):2708-17.

42. An L, Sun MH, Chen F, Li JR. Vitamin D levels in systemic sclerosis patients: A meta-analysis. Drug Design, Development and Therapy. 2017;11:3119-25.

43. Fabisiak N, Fabisiak A, Watala C, Fichna J. Fat-soluble Vitamin Deficiencies and Inflammatory Bowel Disease: Systematic Review and Meta-Analysis. J Clin Gastroenterol. 2017;51(10):878-89.

44. Lu C, Yang J, Yu W, Li D, Xiang Z, Lin Y, et al. Association between 25(OH)D Level, Ultraviolet Exposure, Geographical Location, and Inflammatory Bowel

Disease Activity: A Systematic Review and Meta-Analysis. PLoS ONE [Electronic Resource]. 2015;10(7):e0132036.

45. Wang T, Shan L, Du L, Feng J, Xu Z, Staal WG, et al. Serum concentration of 25-hydroxyvitamin D in autism spectrum disorder: a systematic review and meta-analysis. Eur Child Adolesc Psychiatry. 2016;25(4):341-50.

46. Annweiler C, Milea D, Whitson HE, Cheng CY, Wong TY, Ikram MK, et al. Vitamin D insufficiency and cognitive impairment in Asians: a multi-ethnic population-based study and meta-analysis. J Intern Med. 2016;280(3):300-11.

47. Shen L, Ji HF. Vitamin D deficiency is associated with increased risk of Alzheimer's disease and dementia: evidence from meta-analysis. Nutr J. 2015;14:76.

48. Sommer I, Griebler U, Kien C, Auer S, Klerings I, Hammer R, et al. Vitamin D deficiency as a risk factor for dementia: a systematic review and meta-analysis. BMC geriatr. 2017;17(1):16.

49. Valipour G, Saneei P, Esmaillzadeh A. Serum vitamin D levels in relation to schizophrenia: a systematic review and meta-analysis of observational studies. Journal of Clinical Endocrinology & Metabolism. 2014;99(10):3863-72.

50. Furuya-Kanamori L, Wangdi K, Yakob L, McKenzie SJ, Doi SAR, Clark J, et al. 25-Hydroxyvitamin D Concentrations and Clostridium difficile Infection: A Meta-Analysis. Journal of Parenteral and Enteral Nutrition. 2017;41(5):890-5.

51. Wang YM, Tang CL, Che M, Wang XQ, Li AC. Prognostic value of vitamin D in patients with pneumonia: A systematic review and meta-analysis. Tropical Journal of Pharmaceutical Research. 2017;16(9):2267-73.

52. Riancho-Zarrabeitia L, Cuberia M, Munoz P, Lopez-Hoyos M, Garcia-Canale S, Garcia-Unzueta M, et al. Vitamin D and antiphospholipid syndrome: A retrospective cohort study and meta-analysis. Seminars in Arthritis and Rheumatism. 2017.

53. Qu GB, Wang LL, Tang X, Wu W, Sun YH. The association between vitamin D level and diabetic peripheral neuropathy in patients with type 2 diabetes mellitus: An update systematic review and meta-analysis. Journal of Clinical and Translational Endocrinology. 2017;9:25-31.

54. Derakhshanian H, Shab-Bidar S, Speakman JR, Nadimi H, Djafarian K. Vitamin D and diabetic nephropathy: A systematic review and meta-analysis. Nutrition

(Burbank, Los Angeles County, Calif). 2015;31(10):1189-94.

55. Lv WS, Zhao WJ, Gong SL, Fang DD, Wang B, Fu ZJ, et al. Serum 25-hydroxyvitamin D levels and peripheral neuropathy in patients with type 2 diabetes: a systematic review and meta-analysis. Journal of Endocrinological Investigation. 2015;38(5):513-8.

56. Luo BA, Gao F, Qin LL. The Association between Vitamin D Deficiency and Diabetic Retinopathy in Type 2 Diabetes: A Meta-Analysis of Observational Studies. Nutrients. 2017;9(3):20.

57. Pereira-Santos M, Costa PR, Assis AM, Santos CA, Santos DB. Obesity and vitamin D deficiency: a systematic review and meta-analysis. Obes Rev. 2015;16(4):341-9.

58. Shanmugalingam T, Crawley D, Bosco C, Melvin J, Rohrmann S, Chowdhury S, et al. Obesity and cancer: the role of vitamin D. BMC Cancer. 2014;14:712.

59. Yao Y, Zhu L, He L, Duan Y, Liang W, Nie Z, et al. A meta-analysis of the relationship between vitamin D deficiency and obesity. International Journal of Clinical and Experimental Medicine. 2015;8(9):14977-84.

60. Lucato P, Solmi M, Maggi S, Bertocco A, Bano G, Trevisan C, et al. Low vitamin D levels increase the risk of type 2 diabetes in older adults: A systematic review and meta-analysis. Maturitas. 2017;100:8-15.

61. Hu H, Zhang J, Lu Y, Zhang Z, Qin B, Gao H, et al. Association between Circulating Vitamin D Level and Urolithiasis: A Systematic Review and Meta-Analysis. Nutrients. 2017;9(3):18.

62. Pacheco-Gonzalez RM, Garcia-Marcos L, Morales E. Prenatal vitamin D status and respiratory and allergic outcomes in childhood: A meta-analysis of observational studies. Pediatric Allergy and Immunology. 2018.

63. Jat KR. Vitamin D deficiency and lower respiratory tract infections in children: a systematic review and meta-analysis of observational studies. Trop Doct. 2017;47(1):77-84.

64. Chu J, Gallos I, Tobias A, Tan B, Eapen A, Coomarasamy A. Vitamin D and assisted reproductive treatment outcome: A systematic review and meta-analysis. Human Reproduction. 2018;33(1):65-80.

65. Lv SS, Wang JY, Wang XQ, Wang Y, Xu Y. Serum vitamin D status and in vitro fertilization outcomes: a systematic review and meta-analysis. Archives of Gynecology & Obstetrics. 2016;293(6):1339-45.

66. Bacopoulou F, Kolias E, Efthymiou V, Antonopoulos CN, Charmandari E. Vitamin D predictors in polycystic ovary syndrome: a meta-analysis. European Journal of Clinical Investigation. 2017;47(10):746-55.

67. Guraya SS, Alhussaini KA, Shaqrun FM, Alhazmi BH, Alkabli RS. Correlation of clinical, radiological and serum analysis of hypovitaminosis D with polycystic ovary syndrome: A systematic review and meta-analysis. Journal of Taibah University Medical Sciences. 2017;12(4):277-83.

68. Jia XZ, Wang YM, Zhang N, Guo LN, Zhen XL, Li H, et al. Effect of vitamin D on clinical and biochemical parameters in polycystic ovary syndrome women: A meta-analysis. J Obstet Gynaecol Res. 2015;41(11):1791-802.

69. Amegah AK, Klevor MK, Wagner CL. Maternal Vitamin D insufficiency and risk of adverse pregnancy and birth outcomes: A systematic review and meta-analysis of longitudinal studies. PloS one. 2017;12 (3) (no pagination)(e0173605).

70. Hsiao MY, Hung CY, Chang KV, Han DS, Wang TG. Is serum hypovitaminosis D associated with chronic widespread pain including fibromyalgia? A meta-analysis of observational studies. Pain Physician. 2015;18(5):E877-E87.

71. Makrani AH, Afshari M, Ghajar M, Forooghi Z, Moosazadeh M. Vitamin D and fibromyalgia: A meta-analysis. Korean Journal of Pain. 2017;30(4):250-7.

72. Zadro J, Shirley D, Ferreira M, Carvalho-Silva AP, Lamb SE, Cooper C, et al. Mapping the association between vitamin D and low back pain: A systematic review and Meta-Analysis of observational studies. Pain Physician. 2017;20(7):611-40.

73. Dao D, Sodhi S, Tabasinejad R, Peterson D, Ayeni OR, Bhandari M, et al. Serum 25-Hydroxyvitamin D Levels and Stress Fractures in Military Personnel: A Systematic Review and Meta-analysis. Am J Sports Med. 2015;43(8):2064-72.

74. Annweiler C, Henni S, Walrand S, Montero-Odasso M, Duque G, Duval GT. Vitamin D and walking speed in older adults: Systematic review and meta-analysis. Maturitas. 2017;106:8-25.

75. de Haan K, Groeneveld AB, de Geus HR, Egal M, Struijs A. Vitamin D deficiency

as a risk factor for infection, sepsis and mortality in the critically ill: systematic review and meta-analysis. Crit Care. 2014;18(6):660.

76. Garcia-Alvarez M, Pineda-Tenor D, Jimenez-Sousa MA, Fernandez-Rodriguez A, Guzman-Fulgencio M, Resino S. Relationship of vitamin D status with advanced liver fibrosis and response to hepatitis C virus therapy: a meta-analysis. Hepatology. 2014;60(5):1541-50.

77. Aryan Z, Rezaei N, Camargo CA, Jr. Vitamin D status, aeroallergen sensitization, and allergic rhinitis: A systematic review and meta-analysis. Int Rev Immunol. 2017;36(1):41-53.

78. Annweiler C, Drouet M, Duval GT, Pare PY, Leruez S, Dinomais M, et al. Circulating vitamin D concentration and age-related macular degeneration: Systematic review and meta-analysis. Maturitas. 2016;88:101-12.

79. Khor CC, Do T, Jia H, Nakano M, George R, Abu-Amero K, et al. Genome-wide association study identifies five new susceptibility loci for primary angle closure glaucoma. Nature genetics. 2016;48(5):556-62.

80. Zhu B, Xiao C, Zheng Z. Vitamin D deficiency is associated with the severity of COPD: A systematic review and meta-analysis. International Journal of COPD. 2015;10(1):1907-16.

81. Zhu M, Wang T, Wang C, Ji Y. The association between vitamin D and COPD risk, severity, and exacerbation: an updated systematic review and meta-analysis. International Journal of Copd. 2016;11:2597-607.

82. Jaruvongvanich V, Ahuja W, Sanguankeo A, Wijarnpreecha K, Upala S. Vitamin D and histologic severity of nonalcoholic fatty liver disease: A systematic review and meta-analysis. Digestive and Liver Disease. 2017;49(6):618-22.

83. Zhou J, Huang P, Liu P, Hao Q, Chen S, Dong B, et al. Association of vitamin D deficiency and frailty: A systematic review and meta-analysis. Maturitas. 2016;94:70-6.

84. Zhang YP, Wan YD, Sun TW, Kan QC, Wang LX. Association between vitamin D deficiency and mortality in critically ill adult patients: a meta-analysis of cohort studies. Crit Care. 2014;18(6):684.

85. Wang X, Li W, Zhang Y, Yang Y, Qin G. Association between vitamin D and non-

alcoholic fatty liver disease/non-alcoholic steatohepatitis: Results from a meta-analysis. International Journal of Clinical and Experimental Medicine. 2015;8(10):17221-34.

86. Chowdhury R, Kunutsor S, Vitezova A, Oliver-Williams C, Chowdhury S, Kiefte-de-Jong JC, et al. Vitamin D and risk of cause specific death: systematic review and meta-analysis of observational cohort and randomised intervention studies. BMJ (Clinical research ed). 2014;348:g1903.

87. Neighbors CLP, Noller MW, Song SA, Zaghi S, Neighbors J, Feldman D, et al. Vitamin D and obstructive sleep apnea: a systematic review and meta-analysis. Sleep Medicine. 2018;43:100-8.

88. Lv Z, Qi H, Wang L, Fan X, Han F, Wang H, et al. Vitamin D status and Parkinson's disease: a systematic review and meta-analysis. Neurol Sci. 2014;35(11):1723-30.

89. Rimmelzwaan LM, van Schoor NM, Lips P, Berendse HW, Eekhoff EM. Systematic Review of the Relationship between Vitamin D and Parkinson's Disease. J Parkinsons Dis. 2016;6(1):29-37.

90. Shen L, Ji HF. Associations between Vitamin D Status, Supplementation, Outdoor Work and Risk of Parkinson's Disease: A Meta-Analysis Assessment. Nutrients. 2015;7(6):4817-27.

91. Upala S, Sanguankeo A, Permpalung N. Significant association between vitamin D deficiency and sepsis: a systematic review and meta-analysis. BMC anesthesiol. 2015;15:84.

92. Gaikwad M, Vanlint S, Mittinity M, Moseley GL, Stocks N. Does vitamin D supplementation alleviate chronic nonspecific musculoskeletal pain? A systematic review and meta-analysis. Clinical Rheumatology. 2017;36(5):1201-8.

93. Dewansingh P, Melse-Boonstra A, Krijnen WP, van der Schans CP, Jager-Wittenaar H, van den Heuvel EGHM. Supplemental protein from dairy products increases body weight and vitamin D improves physical performance in older adults: a systematic review and meta-analysis. Nutrition Research. 2018;49:1-22.

94. Diao N, Yang B, Yu F. Effect of vitamin D supplementation on knee osteoarthritis: A systematic review and meta-analysis of randomized clinical trials. Clinical

Biochemistry. 2017;50(18):1312-6.

95. Gao XR, Chen YS, Deng W. The effect of vitamin D supplementation on keen osteoarthritis: A meta-analysis of randomized controlled trials. International Journal of Surgery. 2017;46:14-20.

96. Wu Z, Malihi Z, Stewart AW, Lawes CM, Scragg R. Effect of Vitamin D Supplementation on Pain: A Systematic Review and Meta-analysis. Pain Physician. 2016;19(7):415-27.

97. Yong WC, Sanguankeo A, Upala S. Effect of vitamin D supplementation in chronic widespread pain: a systematic review and meta-analysis. Clinical Rheumatology. 2017;36(12):2825-33.

98. Jiang WL, Gu HB, Zhang YF, Xia QQ, Qi J, Chen JC. Vitamin D Supplementation in the Treatment of Chronic Heart Failure: A Meta-analysis of Randomized Controlled Trials. Clinical Cardiology. 2016;39(1):56-61.

99. Rodriguez AJ, Scott D, Srikanth V, Ebeling P. Effect of vitamin D supplementation on measures of arterial stiffness: a systematic review and meta-analysis of randomized controlled trials. Clinical endocrinology. 2016;84(5):645-57.

100. Upala S, Sanguankeo A, Congrete S, Jaruvongvanich V. Effect of cholecalciferol supplementation on arterial stiffness: a systematic review and meta-analysis. Scand Cardiovasc J. 2016;50(4):230-5.

101. Mann MC, Hobbs AJ, Hemmelgarn BR, Roberts DJ, Ahmed SB, Rabi DM. Effect of oral vitamin D analogs on mortality and cardiovascular outcomes among adults with chronic kidney disease: A meta-analysis. Clinical Kidney Journal. 2015;8(1):41-8.

102. Manousopoulou A, Al-Daghri NM, Garbis SD, Chrousos GP. Vitamin D and cardiovascular risk among adults with obesity: a systematic review and meta-analysis. European Journal of Clinical Investigation. 2015;45(10):1113-26.

103. Hussin AM, Ashor AW, Schoenmakers I, Hill T, Mathers JC, Siervo M. Effects of vitamin D supplementation on endothelial function: a systematic review and meta-analysis of randomised clinical trials. Eur J Nutr. 2017;56(3):1095-104.

104. Joris PJ, Mensink RP. Effects of supplementation with the fat-soluble vitamins E and D on fasting flow-mediated vasodilation in adults: a meta-analysis of

randomized controlled trials. Nutrients. 2015;7(3):1728-43.

105. Mazidi M, Karimi E, Rezaie P, Vatanparast H. The impact of vitamin D supplement intake on vascular endothelial function; a systematic review and meta-analysis of randomized controlled trials. Food and Nutrition Research. 2017;61 (1) (no pagination)(1273574).

106. Stojanovic M, Radenkovic M. Vitamin D versus placebo in improvement of endothelial dysfunction: a meta-analysis of randomized clinical trials. Cardiovasc Ther. 2015;33(3):145-54.

107. Li XH, Feng L, Yang ZH, Liao YH. Effect of active Vitamin D on cardiovascular outcomes in predialysis chronic kidney diseases: A systematic review and meta-analysis. Nephrology. 2015;20(10):706-14.

108. Dinca M, Serban MC, Sahebkar A, Mikhailidis DP, Toth PP, Martin SS, et al. Does vitamin D supplementation alter plasma adipokines concentrations? A systematic review and meta-analysis of randomized controlled trials. Pharmacol Res. 2016;107:360-71.

109. Mirhosseini N, Vatanparast H, Mazidi M, Kimball SM. The Effect of Improved Serum 25-Hydroxyvitamin D Status on Glycemic Control in Diabetic Patients: A Meta-Analysis. Journal of Clinical Endocrinology & Metabolism. 2017;102(9):3097-110.

110. Chandler PD, Wang L, Zhang X, Sesso HD, Moorthy MV, Obi O, et al. Effect of vitamin D supplementation alone or with calcium on adiposity measures: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev. 2015;73(9):577-93.

111. Pathak K, Soares MJ, Calton EK, Zhao Y, Hallett J. Vitamin D supplementation and body weight status: a systematic review and meta-analysis of randomized controlled trials. Obes Rev. 2014;15(6):528-37.

112. Pojsupap S, Iliriani K, Sampaio TZ, O'Hearn K, Kovesi T, Menon K, et al. Efficacy of high-dose vitamin D in pediatric asthma: a systematic review and meta-analysis. J Asthma. 2015;52(4):382-90.

113. Jolliffe DA, Greenberg L, Hooper RL, Griffiths CJ, Camargo CA, Kerley CP, et al. Vitamin D supplementation to prevent asthma exacerbations: a systematic review

and meta-analysis of individual participant data. The Lancet Respiratory Medicine. 2017;5(11):881-90.

114. Riverin BD, Maguire JL, Li P. Vitamin D Supplementation for Childhood Asthma: A Systematic Review and Meta-Analysis. PLoS ONE [Electronic Resource]. 2015;10(8):e0136841.

115. Jamka M, Wozniewicz M, Walkowiak J, Bogdanski P, Jeszka J, Stelmach-Mardas M. The effect of vitamin D supplementation on selected inflammatory biomarkers in obese and overweight subjects: a systematic review with meta-analysis. Eur J Nutr. 2016;55(6):2163-76.

116. Calton EK, Keane KN, Newsholme P, Zhao Y, Soares MJ. The impact of cholecalciferol supplementation on the systemic inflammatory profile: A systematic review and meta-analysis of high-quality randomized controlled trials. European Journal of Clinical Nutrition. 2017;71(8):931-43.

117. Chen N, Wan Z, Han SF, Li BY, Zhang ZL, Qin LQ. Effect of vitamin D supplementation on the level of circulating high-sensitivity C-reactive protein: a meta-analysis of randomized controlled trials. Nutrients. 2014;6(6):2206-16.

118. Langlois PL, Szwec C, D'Aragon F, Heyland DK, Manzanares W. Vitamin D supplementation in the critically ill: A systematic review and meta-analysis. Clinical Nutrition. 2017;18.

119. Putzu A, Belletti A, Cassina T, Clivio S, Monti G, Zangrillo A, et al. Vitamin D and outcomes in adult critically ill patients. A systematic review and meta-analysis of randomized trials. J Crit Care. 2017;38:109-14.

120. Fang F, Ni K, Cai Y, Shang J, Zhang X, Xiong C. Effect of vitamin D supplementation on polycystic ovary syndrome: A systematic review and meta-analysis of randomized controlled trials. Complement Ther Clin Pract. 2017;26:53-60.

121. Wu W, Weng Y, Guo X, Feng L, Xia H, Jiang Z, et al. The Association Between Serum Vitamin D Levels and Age-Related Macular Degeneration: A Systematic Meta-Analytic Review. Invest Ophthalmol Vis Sci. 2016;57(4):2168-77.

122. Cranney A, Horsley T, O'Donnell S, Weiler H, Puil L, Ooi D, et al. Effectiveness and safety of vitamin D in relation to bone health. Evidence report/technology assessment. 2007(158):1-235.

123. Winzenberg T, Powell S, Shaw KA, Jones G. Effects of vitamin D supplementation on bone density in healthy children: systematic review and meta-analysis. BMJ (Clinical research ed). 2011;342:c7254.

124. Reid IR, Bolland MJ, Grey A. Effects of vitamin D supplements on bone mineral density: a systematic review and meta-analysis. Lancet (London, England). 2014;383(9912):146-55.

125. Lai JK, Lucas RM, Clements MS, Roddam AW, Banks E. Hip fracture risk in relation to vitamin D supplementation and serum 25-hydroxyvitamin D levels: a systematic review and meta-analysis of randomised controlled trials and observational studies. BMC public health. 2010;10:331.

126. Bischoff-Ferrari HA, Willett WC, Wong JB, Stuck AE, Staehelin HB, Orav EJ, et al. Prevention of nonvertebral fractures with oral vitamin D and dose dependency: a meta-analysis of randomized controlled trials. Archives of internal medicine. 2009;169(6):551-61.

127. Patient level pooled analysis of 68 500 patients from seven major vitamin D fracture trials in US and Europe. BMJ (Clinical research ed). 2010;340:b5463.

128. Bischoff-Ferrari HA, Willett WC, Orav EJ, Lips P, Meunier PJ, Lyons RA, et al. A pooled analysis of vitamin D dose requirements for fracture prevention. The New England journal of medicine. 2012;367(1):40-9.

129. Avenell A, Gillespie WJ, Gillespie LD, O'Connell D. Vitamin D and vitamin D analogues for preventing fractures associated with involutional and post-menopausal osteoporosis. The Cochrane database of systematic reviews. 2009(2):Cd000227.

130. Bergman GJ, Fan T, McFetridge JT, Sen SS. Efficacy of vitamin D3 supplementation in preventing fractures in elderly women: a meta-analysis. Current medical research and opinion. 2010;26(5):1193-201.

131. Bischoff-Ferrari HA, Willett WC, Wong JB, Giovannucci E, Dietrich T, Dawson-Hughes B. Fracture prevention with vitamin D supplementation: a meta-analysis of randomized controlled trials. Jama. 2005;293(18):2257-64.

132. Chung M, Lee J, Terasawa T, Lau J, Trikalinos TA. Vitamin D with or without calcium supplementation for prevention of cancer and fractures: an updated meta-

analysis for the U.S. Preventive Services Task Force. Annals of internal medicine. 2011;155(12):827-38.

133. Tang BM, Eslick GD, Nowson C, Smith C, Bensoussan A. Use of calcium or calcium in combination with vitamin D supplementation to prevent fractures and bone loss in people aged 50 years and older: a meta-analysis. Lancet (London, England). 2007;370(9588):657-66.

134. Bischoff-Ferrari HA, Dawson-Hughes B, Staehelin HB, Orav JE, Stuck AE, Theiler R, et al. Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials. BMJ (Clinical research ed). 2009;339:b3692.

135. Bischoff-Ferrari HA, Dawson-Hughes B, Willett WC, Staehelin HB, Bazemore MG, Zee RY, et al. Effect of Vitamin D on falls: a meta-analysis. Jama. 2004;291(16):1999-2006.

136. Lee JE, Li H, Chan AT, Hollis BW, Lee IM, Stampfer MJ, et al. Circulating levels of vitamin D and colon and rectal cancer: the Physicians' Health Study and a meta-analysis of prospective studies. Cancer prevention research (Philadelphia, Pa). 2011;4(5):735-43.

137. Yin L, Grandi N, Raum E, Haug U, Arndt V, Brenner H. Meta-analysis: Serum vitamin D and colorectal adenoma risk. Preventive medicine. 2011;53(1-2):10-6.

138. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, et al. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. International journal of cancer. 2011;128(6):1414-24.

139. Gilbert R, Martin RM, Beynon R, Harris R, Savovic J, Zuccolo L, et al. Associations of circulating and dietary vitamin D with prostate cancer risk: a systematic review and dose-response meta-analysis. Cancer causes & control : CCC. 2011;22(3):319-40.

140. Yin L, Raum E, Haug U, Arndt V, Brenner H. Meta-analysis of longitudinal studies: Serum vitamin D and prostate cancer risk. Cancer epidemiology. 2009;33(6):435-45.

141. Chen P, Li M, Gu X, Liu Y, Li X, Li C, et al. Higher blood 25(OH)D level may reduce the breast cancer risk: evidence from a Chinese population based case-control study and meta-analysis of the observational studies. PloS one. 2013;8(1):e49312.

142. Bauer SR, Hankinson SE, Bertone-Johnson ER, Ding EL. Plasma vitamin D levels, menopause, and risk of breast cancer: dose-response meta-analysis of prospective studies. Medicine. 2013;92(3):123-31.

143. Burgaz A, Orsini N, Larsson SC, Wolk A. Blood 25-hydroxyvitamin D concentration and hypertension: a meta-analysis. Journal of hypertension. 2011;29(4):636-45.

144. Witham MD, Nadir MA, Struthers AD. Effect of vitamin D on blood pressure: a systematic review and meta-analysis. Journal of hypertension. 2009;27(10):1948-54.

145. Wu SH, Ho SC, Zhong L. Effects of vitamin D supplementation on blood pressure. Southern medical journal. 2010;103(8):729-37.

146. Brondum-Jacobsen P, Nordestgaard BG, Schnohr P, Benn M. 25-hydroxyvitamin D and symptomatic ischemic stroke: an original study and meta-analysis. Annals of neurology. 2013;73(1):38-47.

147. Sun Q, Pan A, Hu FB, Manson JE, Rexrode KM. 25-Hydroxyvitamin D levels and the risk of stroke: a prospective study and meta-analysis. Stroke; a journal of cerebral circulation. 2012;43(6):1470-7.

148. Wang L, Song Y, Manson JE, Pilz S, Marz W, Michaelsson K, et al. Circulating 25-hydroxy-vitamin D and risk of cardiovascular disease: a meta-analysis of prospective studies. Circulation Cardiovascular quality and outcomes. 2012;5(6):819-29.

149. Brondum-Jacobsen P, Benn M, Jensen GB, Nordestgaard BG. 25-hydroxyvitamin d levels and risk of ischemic heart disease, myocardial infarction, and early death: population-based study and meta-analyses of 18 and 17 studies. Arteriosclerosis, thrombosis, and vascular biology. 2012;32(11):2794-802.

150. Myung SK, Ju W, Cho B, Oh SW, Park SM, Koo BK, et al. Efficacy of vitamin and antioxidant supplements in prevention of cardiovascular disease: systematic review and meta-analysis of randomised controlled trials. BMJ (Clinical research ed).

2013;346:f10.

151. Forouhi NG, Ye Z, Rickard AP, Khaw KT, Luben R, Langenberg C, et al. Circulating 25-hydroxyvitamin D concentration and the risk of type 2 diabetes: results from the European Prospective Investigation into Cancer (EPIC)-Norfolk cohort and updated meta-analysis of prospective studies. Diabetologia. 2012;55(8):2173-82.

152. Afzal S, Bojesen SE, Nordestgaard BG. Low 25-hydroxyvitamin D and risk of type 2 diabetes: a prospective cohort study and metaanalysis. Clinical chemistry. 2013;59(2):381-91.

153. George PS, Pearson ER, Witham MD. Effect of vitamin D supplementation on glycaemic control and insulin resistance: a systematic review and meta-analysis. Diabetic medicine : a journal of the British Diabetic Association. 2012;29(8):e142-50.

154. Balion C, Griffith LE, Strifler L, Henderson M, Patterson C, Heckman G, et al. Vitamin D, cognition, and dementia: a systematic review and meta-analysis. Neurology. 2012;79(13):1397-405.

155. Etgen T, Sander D, Bickel H, Sander K, Forstl H. Vitamin D deficiency, cognitive impairment and dementia: a systematic review and meta-analysis. Dementia and geriatric cognitive disorders. 2012;33(5):297-305.

156. Anglin RE, Samaan Z, Walter SD, McDonald SD. Vitamin D deficiency and depression in adults: systematic review and meta-analysis. The British journal of psychiatry : the journal of mental science. 2013;202:100-7.

157. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, et al. Genome-wide association study of circulating vitamin D levels. Human molecular genetics. 2010;19(13):2739-45.

158. Jiang X, O'Reilly PF, Aschard H, Hsu YH, Richards JB, Dupuis J, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. Nature communications. 2018;9(1):260.

159. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet (London, England). 2010;376(9736):180-8.

160. Sinotte M, Diorio C, Berube S, Pollak M, Brisson J. Genetic polymorphisms

of the vitamin D binding protein and plasma concentrations of 25-hydroxyvitamin D in premenopausal women. The American journal of clinical nutrition. 2009;89(2):634-40.

161. Speeckaert M, Huang G, Delanghe JR, Taes YE. Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. Clin Chim Acta. 2006;372(1-2):33-42.

162. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(20):7711-5.

163. Wassif CA, Maslen C, Kachilele-Linjewile S, Lin D, Linck LM, Connor WE, et al. Mutations in the human sterol delta7-reductase gene at 11q12-13 cause Smith-Lemli-Opitz syndrome. American journal of human genetics. 1998;63(1):55-62.

164. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nature genetics. 2004;36(1):40-5.

165. Ahn J, Albanes D, Berndt SI, Peters U, Chatterjee N, Freedman ND, et al. Vitamin D-related genes, serum vitamin D concentrations and prostate cancer risk. Carcinogenesis. 2009;30(5):769-76.

166. Greenland S. An introduction To instrumental variables for epidemiologists. International journal of epidemiology. 2000;29(6):1102.

167. Thomas DC, Conti DV. Commentary: the concept of 'Mendelian Randomization'. International journal of epidemiology. 2004;33(1):21-5.

168. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology (Cambridge, Mass). 2014;25(3):427-35.

169. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008;27(8):1133-63.

170. Abbasi A. Mendelian randomization studies of biomarkers and type 2 diabetes. Endocrine Connections. 2015;4(4):249-60.

171. Afzal S, Brondum-Jacobsen P, Bojesen SE, Nordestgaard BG. Vitamin D concentration, obesity, and risk of diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2014;2(4):298-306.

172. Husemoen LL, Skaaby T, Martinussen T, Jorgensen T, Thuesen BH, Kistorp C, et al. Investigating the causal effect of vitamin D on serum adiponectin using a Mendelian randomization approach. European Journal of Clinical Nutrition. 2014;68(2):189-95.

173. Leong A, Rehman W, Dastani Z, Greenwood C, Timpson N, Langsetmo L, et al. The causal effect of vitamin D binding protein (DBP) levels on calcemic and cardiometabolic diseases: a Mendelian randomization study. PLoS Med. 2014;11(10):e1001751.

174. Ye Z, Sharp SJ, Burgess S, Scott RA, Imamura F, Langenberg C, et al. Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: A mendelian randomisation study. The Lancet Diabetes and Endocrinology. 2015;3(1):35-42.

175. Brondum-Jacobsen P, Benn M, Afzal S, Nordestgaard BG. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: a Mendelian randomization study. International journal of epidemiology. 2015;44(2):651-61.

176. Kunutsor SK, Burgess S, Munroe PB, Khan H. Vitamin D and high blood pressure: causal association or epiphenomenon? European Journal of Epidemiology. 2014;29(1):1-14.

177. Manousaki D, Mokry LE, Ross S, Goltzman D, Brent Richards J. Mendelian Randomization Studies Do Not Support a Role for Vitamin D in Coronary Artery Disease. Circulation: Cardiovascular Genetics. 2016;9(4):349-56.

178. Skaaby T, Husemoen LL, Martinussen T, Thyssen JP, Melgaard M, Thuesen BH, et al. Vitamin D status, filaggrin genotype, and cardiovascular risk factors: a Mendelian randomization approach. PLoS ONE [Electronic Resource]. 2013;8(2):e57647.

179. Vimaleswaran KS, Cavadino A, Berry DJ, LifeLines Cohort Study i, Jorde R, Dieffenbach AK, et al. Association of vitamin D status with arterial blood pressure and

hypertension risk: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2014;2(9):719-29.

180. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Goltzman D, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study.[Erratum appears in PLoS Med. 2016 Mar;13(3):e1001981 Note: Goltzman, David [added]; PMID: 26963823]. PLoS Med. 2015;12(8):e1001866.

181. Rhead B, Baarnhielm M, Gianfrancesco M, Mok A, Shao X, Quach H, et al. Mendelian randomization shows a causal effect of low Vitamin D on multiple sclerosis risk. Neurology: Genetics. 2016;2 (5) (no pagination)(e97).

182. Viatte S, Yarwood A, McAllister K, Al-Mudhaffer S, Fu B, Flynn E, et al. The role of genetic polymorphisms regulating vitamin D levels in rheumatoid arthritis outcome: a Mendelian randomisation approach. Annals of the Rheumatic Diseases. 2014;73(7):1430-3.

183. Yarwood A, Viatte S, Plant D, Morgan AW, Isaacs J, Wilson AG, et al. Testing the role of vitamin D in response to antitumour necrosis factor alpha therapy in a UK cohort: a Mendelian randomisation approach. Annals of the Rheumatic Diseases. 2014;73(5):938-40.

184. Jorde R, Mathiesen EB, Rogne S, Wilsgaard T, Kjaergaard M, Grimnes G, et al. Vitamin D and cognitive function: The Tromso Study. Journal of the Neurological Sciences. 2015;355(1-2):155-61.

185. Kueider AM, Tanaka T, An Y, Kitner-Triolo MH, Palchamy E, Ferrucci L, et al. State- and trait-dependent associations of vitamin-D with brain function during aging. Neurobiology of Aging. 2016;39:38-45.

186. Mokry LE, Ross S, Morris JA, Manousaki D, Forgetta V, Richards JB. Genetically decreased Vitamin D and risk of Alzheimer disease. Neurology. 2016;87(24):2567-74.

187. Afzal S, Brondum-Jacobsen P, Bojesen SE, Nordestgaard BG. Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts. BMJ (Clinical research ed). 2014;349:g6330.

188. Davies JR, Field S, Randerson-Moor J, Harland M, Kumar R, Anic GM, et al. An inherited variant in the gene coding for vitamin D-binding protein and survival

from cutaneous melanoma: a BioGenoMEL study. Pigment Cell Melanoma Res. 2014;27(2):234-43.

189. Trummer O, Pilz S, Hoffmann MM, Winkelmann BR, Boehm BO, Marz W, et al. Vitamin D and mortality: a Mendelian randomization study. Clinical chemistry. 2013;59(5):793-7.

190. Theodoratou E, Palmer T, Zgaga L, Farrington SM, McKeigue P, Din FV, et al. Instrumental variable estimation of the causal effect of plasma 25-hydroxy-vitamin D on colorectal cancer risk: a mendelian randomization analysis. PLoS ONE [Electronic Resource]. 2012;7(6):e37662.

191. Ong JS, Cuellar-Partida G, Lu Y, Fasching PA, Hein A, Burghaus S, et al. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. International journal of epidemiology. 2016;45(5):1619-30.

192. Dudding T, Thomas SJ, Duncan K, Lawlor DA, Timpson NJ. Re-Examining the Association between Vitamin D and Childhood Caries. PLoS ONE [Electronic Resource]. 2015;10(12):e0143769.

193. Li SS, Gao LH, Zhang XY, He JW, Fu WZ, Liu YJ, et al. Genetically Low Vitamin D Levels, Bone Mineral Density, and Bone Metabolism Markers: a Mendelian Randomisation Study. Sci Rep. 2016;6:33202.

194. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, et al. Genetic evidence for causal relationships between maternal obesity-related traits and birth weight. JAMA - Journal of the American Medical Association. 2016;315(11):1129-40.

195. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med. 2013;10(2):e1001383.

196. Hysinger EB, Roizen JD, Mentch FD, Vazquez L, Connolly JJ, Bradfield JP, et al. Mendelian randomization analysis demonstrates that low vitamin D is unlikely causative for pediatric asthma. Journal of Allergy and Clinical Immunology. 2016;138(6):1747-9.e4.

197. Liefaard MC, Ligthart S, Vitezova A, Hofman A, Uitterlinden AG, Kiefte-de Jong JC, et al. Vitamin D and C-Reactive Protein: A Mendelian Randomization Study.

PLoS ONE [Electronic Resource]. 2015;10(7):e0131740.

198. Ooi EM, Afzal S, Nordestgaard BG. Elevated remnant cholesterol in 25-hydroxyvitamin D deficiency in the general population: Mendelian randomization study. Circ Cardiovasc Genet. 2014;7(5):650-8.

199. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics (Oxford, England). 2010;26(9):1205-10.

200. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. American journal of human genetics. 2011;89(4):529-42.

201. Pathak J, Kiefer RC, Bielinski SJ, Chute CG. Mining the human phenome using semantic web technologies: a case study for Type 2 Diabetes. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;2012:699-708.

202. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature biotechnology. 2013;31(12):1102-10.

203. Hebbring SJ, Schrodi SJ, Ye Z, Zhou Z, Page D, Brilliant MH. A PheWAS approach in studying HLA-DRB1*1501. Genes and immunity. 2013;14(3):187-91.

204. Liao KP, Kurreeman F, Li G, Duclos G, Murphy S, Guzman R, et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls. Arthritis and rheumatism. 2013;65(3):571-81.

205. Neuraz A, Chouchana L, Malamut G, Le Beller C, Roche D, Beaune P, et al. Phenome-wide association studies on a quantitative trait: application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. PLoS computational biology. 2013;9(12):e1003405.

206. Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers

of arrhythmia risk. Circulation. 2013;127(13):1377-85.

207. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics (Oxford, England). 2014;30(16):2375-6.

208. Cronin RM, Field JR, Bradford Y, Shaffer CM, Carroll RJ, Mosley JD, et al. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Frontiers in genetics. 2014;5:250.

209. Mitchell SL, Hall JB, Goodloe RJ, Boston J, Farber-Eger E, Pendergrass SA, et al. Investigating the relationship between mitochondrial genetic variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies. BioData mining. 2014;7:6.

210. Namjou B, Marsolo K, Caroll RJ, Denny JC, Ritchie MD, Verma SS, et al. Phenome-wide association study (PheWAS) in EMR-linked pediatric cohorts, genetically links PLCL1 to speech language development and IL5-IL13 to Eosinophilic Esophagitis. Frontiers in genetics. 2014;5:401.

211.Shameer K, Denny JC, Ding K, Jouni H, Crosslin DR, de Andrade M, et al. A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. Human genetics. 2014;133(1):95-109.

212. Boland MR, Shahn Z, Madigan D, Hripcsak G, Tatonetti NP. Birth month affects lifetime disease risk: A phenome-wide method. Journal of the American Medical Informatics Association. 2015;22(5):1042-53.

213. Diogo D, Bastarache L, Liao KP, Graham RR, Fulton RS, Greenberg JD, et al. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. PloS one. 2015;10(4):e0122271.

214. Namjou B, Marsolo K, Lingren T, Ritchie MD, Verma SS, Cobb BL, et al. A GWAS Study on Liver Function Test Using eMERGE Network Participants. PLoS ONE [Electronic Resource]. 2015;10(9):e0138677.

215. Ye Z, Mayer J, Ivacic L, Zhou Z, He M, Schrodi SJ, et al. Phenome-wide

association studies (PheWASs) for functional variants. European journal of human genetics : EJHG. 2015;23(4):523-9.

216. Hebbring SJ, Rastegar-Mojarad M, Ye Z, Mayer J, Jacobson C, Lin S. Application of clinical text data for phenome-wide association studies (PheWASs). Bioinformatics (Oxford, England). 2015;31(12):1981-7.

217. Li L, Boland MR, Miotto R, Tatonetti NP, Dudley JT. Replicating Cardiovascular Condition-Birth Month Associations. Sci. 2016;6:33166.

218. Liu J, Ye Z, Mayer JG, Hoch BA, Green C, Rolak L, et al. Phenome-wide association study maps new diseases to the human major histocompatibility complex region. J Med Genet. 2016;53(10):681-9.

219. Millwood IY, Bennett DA, Walters RG, Clarke R, Waterworth D, Johnson T, et al. A phenome-wide association study of a lipoprotein-associated phospholipase A2 loss-of-function variant in 90 000 Chinese adults. International journal of epidemiology. 2016;45(5):1588-99.

220. Mosley JD, Witte JS, Larkin EK, Bastarache L, Shaffer CM, Karnes JH, et al. Identifying genetically driven clinical phenotypes using linear mixed models. Nature communications. 2016;7:11433.

221. Oetjens MT, Bush WS, Denny JC, Birdwell K, Kodaman N, Verma A, et al. Evidence for extensive pleiotropy among pharmacogenes. Pharmacogenomics. 2016;17(8):853-66.

222. Verma A, Basile AO, Bradford Y, Kuivaniemi H, Tromp G, Carey D, et al. Phenome-Wide Association Study to Explore Relationships between Immune System Related Genetic Loci and Complex Traits and Diseases. PLoS ONE [Electronic Resource]. 2016;11(8):e0160573.

223. Verma A, Leader JB, Verma SS, Frase A, Wallace J, Dudek S, et al. Integrating Clinical Laboratory Measures and Icd-9 Code Diagnoses in Phenome-Wide Association Studies. Pacific Symposium on Biocomputing. 2016;21:168-79.

224. Verma A, Verma SS, Pendergrass SA, Crawford DC, Crosslin DR, Kuivaniemi H, et al. eMERGE Phenome-Wide Association Study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants. BMC Med Genomics. 2016;9 Suppl 1:32.

225. Verma SS, Lucas AM, Lavage DR, Leader JB, Metpally R, Krishnamurthy S, et al. Identifying Genetic Associations with Variability in Metabolic Health and Blood Count Laboratory Values: Diving into the Quantitative Traits by Leveraging Longitudinal Data from an Ehr. Pacific Symposium on Biocomputing. 2016;Pacific Symposium on Biocomputing. 22:533-44.

226. Cortes A, Dendrou CA, Motyer A, Jostins L, Vukcevic D, Dilthey A, et al. Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. Nature genetics. 2017;49(9):1311-8.

227. Doss J, Mo H, Carroll RJ, Crofford LJ, Denny JC. Phenome-Wide Association Study of Rheumatoid Arthritis Subgroups Identifies Association Between Seronegative Disease and Fibromyalgia. Arthritis rheumatol. 2017;69(2):291-300.

228. Jannot AS, Burgun A, Thervet E, Pallet N. The Diagnosis-Wide Landscape of Hospital-Acquired AKI. Clinical Journal of The American Society of Nephrology: CJASN. 2017;12(6):874-84.

229. Karnes JH, Bastarache L, Shaffer CM, Gaudieri S, Xu Y, Glazer AM, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. Sci Transl Med. 2017;9(389):10.

230. Liao KP, Sparks JA, Hejblum BP, Kuo IH, Cui J, Lahey LJ, et al. Phenome-Wide Association Study of Autoantibodies to Citrullinated and Noncitrullinated Epitopes in Rheumatoid Arthritis. Arthritis rheumatol. 2017;69(4):742-9.

231. Liu J, Zhao R, Ye Z, Frey AJ, Schriver ER, Snyder NW, et al. Relationship of SULT1A1 copy number variation with estrogen metabolism and human health. Journal of Steroid Biochemistry & Molecular Biology. 2017;174:169-75.

232. Robinson JR, Kennedy VE, Doss Y, Bastarache L, Denny J, Warner JL. Defining the complex phenotype of severe systemic loxoscelism using a large electronic health record cohort. PLoS ONE [Electronic Resource]. 2017;12(4):e0174941.

233. Pendergrass SA, Brown-Gentry K, Dudek S, Frase A, Torstenson ES, Goodloe R, et al. Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS genetics. 2013;9(1):e1003087.

234. Hall MA, Verma A, Brown-Gentry KD, Goodloe R, Boston J, Wilson S, et al. Detection of pleiotropy through a Phenome-wide association study (PheWAS) of epidemiologic data as part of the Environmental Architecture for Genes Linked to Environment (EAGLE) study. PLoS genetics. 2014;10(12):e1004678.

235. Millard LA, Davies NM, Timpson NJ, Tilling K, Flach PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Sci. 2015;5:16645.

236. Moore CB, Verma A, Pendergrass S, Verma SS, Johnson DH, Daar ES, et al. Phenome-wide Association Study Relating Pretreatment Laboratory Parameters With Human Genetic Variants in AIDS Clinical Trials Group Protocols. Open forum infectious diseases. 2015;2(1):ofu113.

237. Roesch SL, Styer AM, Wood GC, Kosak Z, Seiler J, Benotti P, et al. Perturbations of fibroblast growth factors 19 and 21 in type 2 diabetes. PloS one. 2015;10(2):e0116928.

238. Karaca S, Civelek E, Karaca M, Sahiner UM, Ozgul RK, Kocabas CN, et al. Allergy-specific Phenome-Wide Association Study for Immunogenes in Turkish Children. Sci. 2016;6:33152.

239. Krapohl E, Euesden J, Zabaneh D, Pingault JB, Rimfeld K, von Stumm S, et al. Phenome-wide analysis of genome-wide polygenic scores. Mol Psychiatry. 2016;21(9):1188-93.

240. Polimanti R, Kranzler HR, Gelernter J. Phenome-Wide Association Study for Alcohol and Nicotine Risk Alleles in 26394 Women. Neuropsychopharmacology. 2016;41(11):2688-96.

241. Ehm MG, Aponte JL, Chiano MN, Yerges-Armstrong LM, Johnson T, Barker JN, et al. Phenome-wide association study using research participants' self-reported data provides insight into the Th17 and IL-17 pathway. PLoS ONE [Electronic Resource]. 2017;12(11):e0186405.

242. Polimanti R, Jensen KP, Gelernter J. Phenome-wide association study for CYP2A6 alleles: rs113288603 is associated with hearing loss symptoms in elderly smokers. Sci. 2017;7(1):1034.

243. Verma A, Bradford Y, Verma SS, Pendergrass SA, Daar ES, Venuto C, et al.

Multiphenotype association study of patients randomized to initiate antiretroviral regimens in AIDS Clinical Trials Group protocol A5202. Pharmacogenetics and Genomics. 2017;27(3):101-11.

244. AHRQ Clinical Classifications Software (CCS) for ICD-9-CM Fact Sheet. [Internet]. Available from: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccsfactsheet.jsp.

245. Burwinkel B, Shanmugam KS, Hemminki K, Meindl A, Schmutzler RK, Sutter C, et al. Transcription factor 7-like 2 (TCF7L2) variant is associated with familial breast cancer risk: a case-control study. BMC Cancer. 2006;6:268.

246. Agalliu I, Suuriniemi M, Prokunina-Olsson L, Johanneson B, Collins FS, Stanford JL, et al. Evaluation of a variant in the transcription factor 7-like 2 (TCF7L2) gene and prostate cancer risk in a population-based study. The Prostate. 2008;68(7):740-7.

247. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. Nature reviews Genetics. 2016;17(3):129-45.

248. Qin C, Lv J, Guo Y, Bian Z, Si J, Yang L, et al. Associations of egg consumption with cardiovascular disease in a cohort study of 0.5 million Chinese adults. Heart (British Cardiac Society). 2018.

249. Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. Genetics. 2015;200(4):1285-95.

250. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. Journal of clinical epidemiology. 2016;70:214-23.

251. Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. Immunology. 2014;141(2):157-65.

252. Fortier I, Doiron D, Little J, Ferretti V, L'Heureux F, Stolk RP, et al. Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies. International journal of epidemiology. 2011;40(5):1314-28.

253. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics. 2001;29(4):1165-88.

254. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD. Visually integrating and exploring high throughput Phenome-Wide Association Study (PheWAS) results using PheWAS-View. BioData mining. 2012;5(1):5.

255. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

256. Biobank U. [Available from: http://biobank.ctsu.ox.ac.uk/crystal/docs/UKgrid.pdf.

257. Biobank U. HES data [Available from: http://biobank.ctsu.ox.ac.uk/showcase/docs/HospitalEpisodeStatistics.pdf.

258. Biobank U. UK Biobank cancer register [Available from: http://biobank.ctsu.ox.ac.uk/crystal/docs/CancerLinkage.pdf.

259. Biobank U. UK Biobank death register [Available from: http://biobank.ctsu.ox.ac.uk/crystal/docs/DeathLinkage.pdf.

260. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. The Lancet Respiratory medicine. 2015;3(10):769-81.

261. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

262. Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, et al. Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. American journal of human genetics. 2016;98(3):456-72.

263. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics (Oxford, England). 2010;26(22):2867-73.

264. Verma A, Bradford Y, Dudek S, Verma SS, Pendergrass SA, Ritchie MD. A simulation study investigating power estimates in Phenome-Wide Association Studies.

bioRxiv. 2017.

265. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. International journal of epidemiology. 2014;43(3):922-9.

266. Timofeeva MN, Kinnersley B, Farrington SM, Whiffin N, Palles C, Svinti V, et al. Recurrent Coding Sequence Variation Explains Only A Small Fraction of the Genetic Architecture of Colorectal Cancer. Sci Rep. 2015;5:16286.

267. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

268. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.

269. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat Med. 2015;34(21):2926-40.

270. Husemoen L, Skaaby T, Martinussen T, Jorgensen T, Thuesen B, Kistorp C, et al. Investigating the causal effect of vitamin D on serum adiponectin using a Mendelian randomization approach. European journal of epidemiology. 2013;1):S77-S8.

271. Trummer O, Pilz S, Hoffmann M, Winkelmann B, Bohm B, Marz W, et al. Vitamin D and mortality: A mendelian randomization study. Austrian Journal of Clinical Endocrinology and Metabolism. 2012;5:16.

272. Yarwood A, Martin P, Bowes J, Lunt M, Worthington J, Barton A, et al. Enrichment of vitamin D response elements in RA-associated loci supports a role for vitamin D in the pathogenesis of RA. Genes and immunity. 2013;14(5):325-9.

273. Xie Y. Population heterogeneity and causal inference. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(16):6262-8.

274. Meng XR, Song JY, Ma J, Liu FH, Shang XR, Guo XJ, et al. Association study of childhood obesity with eight genetic variants recently identified by genome-wide association studies. Pediatric research. 2014;76(3):310-5.

275. Ioannidis JP. Why most discovered true associations are inflated. Epidemiology. 2008:640-8.

276. Willis BH, Riley RD. Measuring the statistical validity of summary meta-analysis and meta-regression results for use in clinical practice. Stat Med. 2017;36(21):3283-301.

277. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.

278. Kontopantelis E, Springate DA, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. PloS one. 2013;8(7):e69930.

279. Kontopantelis E, Reeves D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. Statistical methods in medical research. 2012;21(4):409-26.

280. Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, Cotsapas C, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nature genetics. 2013;45(11):1353-60.

281. Bennett SN, Caporaso N, Fitzpatrick AL, Agrawal A, Barnes K, Boyd HA, et al. Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience. Genetic epidemiology. 2011;35(3):159-73.

282. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. 2009.

283. Woodfield R, Sudlow CL. Accuracy of patient self-report of stroke: a systematic review from the UK Biobank Stroke Outcomes Group. PloS one. 2015;10(9):e0137538.

284. Nevitt MC, Cummings SR, Browner WS, Seeley DG, Cauley JA, Vogt SR, et al. The accuracy of self-report of fractures in elderly women: evidence from a prospective study. American Journal of Epidemiology. 1992;135(5):490-9.

285. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. Journal of the American Medical Informatics Association. 2016;23(5):1007-15.

286. Li L, Chase HS, Patel CO, Friedman C, Weng C, editors. Comparing ICD9-encoded diagnoses and NLP-processed discharge summaries for clinical trials pre-screening: a case study. AMIA Annual Symposium Proceedings; 2008: American Medical Informatics Association.

287. Baus A, Hendryx M, Pollard C. Identifying patients with hypertension: a case for auditing electronic health record data. Perspectives in health information management/AHIMA, American Health Information Management Association. 2012;9(Spring).

288. Wei W-Q, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome medicine. 2015;7(1):41.

289. Reid IR, Horne AM, Mihov B, Gamble GD, Al-Abuwsi F, Singh M, et al. Effect of monthly high-dose vitamin D on bone density in community-dwelling older adults substudy of a randomized controlled trial. J Intern Med. 2017;282(5):452-60.

290. Macdonald HM, Reid IR, Gamble GD, Fraser WD, Tang JC, Wood AD. 25-Hydroxyvitamin D Threshold for the Effects of Vitamin D Supplements on Bone Density Secondary Analysis of a Randomized Controlled Trial. Journal of Bone and Mineral Research. 2018.

291. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology. 2012;1(3):123-6.

292. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. American journal of epidemiology. 2017;186(9):1026-34.

293. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. International journal of epidemiology. 2009;39(2):417-20.

294. Pendergrass S, Ritchie MD. Phenome-wide association studies: leveraging comprehensive phenotypic and genotypic data for discovery. Current genetic medicine reports. 2015;3(2):92-100.

295. Pendergrass SA, Verma A, Okula A, Hall MA, Crawford DC, Ritchie MD.

Phenome-Wide Association Studies: Embracing Complexity for Discovery. Hum Hered. 2015;79(3-4):111-23.

296. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. Annu Rev Genomics Hum Genet. 2016;17:353-73.

297. Roden DM. Phenome-wide association studies: a new method for functional genomics in humans. J Physiol (Lond). 2017;595(12):4109-15.

298. Wei Z, Zhang J, Yu X. Maternal vitamin D status and childhood asthma, wheeze, and eczema: A systematic review and meta-analysis. Pediatr Allergy Immunol. 2016;27(6):612-9.

299. Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei WQ, Gifford A, et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. Ann Rheum Dis. 2018;77(7):1039-47.

300. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. Journal of clinical epidemiology. 2014;67(8):850-7.

301. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B (Methodological). 1995:289-300.

302. Cortes A, Dendrou CA, Motyer A, Jostins L, Vukcevic D, Dilthey A, et al. Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. Nature genetics. 2017;49(9):1311.

303. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. International journal of epidemiology. 2011;40(3):755-64.

304. Karohl C, Su S, Kumari M, Tangpricha V, Veledar E, Vaccarino V, et al. Heritability and seasonal variability of vitamin D concentrations in male twins. The American journal of clinical nutrition. 2010;92(6):1393-8.

305. Orton SM, Morris AP, Herrera BM, Ramagopalan SV, Lincoln MR, Chao MJ, et al. Evidence for genetic regulation of vitamin D status in twins with multiple sclerosis. The American journal of clinical nutrition. 2008;88(2):441-7.

306. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Statistical methods in medical research. 2017;26(5):2333-55.

307. Palmer TM, Holmes MV, Keating BJ, Sheehan NA. Correcting the Standard Errors of 2-Stage Residual Inclusion Estimators for Mendelian Randomization Studies. American journal of epidemiology. 2017;186(9):1104-14.

308. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. European journal of epidemiology. 2017;32(5):377-89.

309. Berry DJ, Vimaleswaran KS, Whittaker JC, Hingorani AD, Hypponen E. Evaluation of genetic markers as instruments for Mendelian randomization studies on vitamin D. PloS one. 2012;7(5):e37465.

310. Stefansson TS, McDonald BA, Willi Y. The influence of genetic drift and selection on quantitative traits in a plant pathogenic fungus. PloS one. 2014;9(11):e112523.

311.Berry DJ, Vimaleswaran KS, Whittaker JC, Hingorani AD, Hypponen E. Evaluation of genetic markers as instruments for Mendelian randomization studies on vitamin D. PLoS ONE [Electronic Resource]. 2012;7(5):e37465.

312. Liebeskind DS. Nephrotic syndrome. Handbook of clinical neurology. 2014;119:405-15.

313. Paccaud JP, Reith W, Carpentier JL, Ravazzola M, Amherdt M, Schekman R, et al. Cloning and functional characterization of mammalian homologues of the COPII component Sec23. Molecular biology of the cell. 1996;7(10):1535-46.

314. Ferraro PM, Minucci A, Primiano A, De Paolis E, Gervasoni J, Persichilli S, et al. A novel CYP24A1 genotype associated to a clinical picture of hypercalcemia, nephrolithiasis and low bone mass. Urolithiasis. 2017;45(3):291-4.

315. Dinour D, Beckerman P, Ganon L, Tordjman K, Eisenstein Z, Holtzman EJ. Loss-of-function mutations of CYP24A1, the vitamin D 24-hydroxylase gene, cause long-standing hypercalciuric nephrolithiasis and nephrocalcinosis. The Journal of urology. 2013;190(2):552-7.

316. Petkovich M, Jones G. CYP24A1 and kidney disease. Current opinion in

nephrology and hypertension. 2011;20(4):337-44.

317. Ye Z, Sharp SJ, Burgess S, Scott RA, Imamura F, InterAct C, et al. Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: a mendelian randomisation study.[Erratum appears in Lancet Diabetes Endocrinol. 2015 Jan;3(1):e1]. Lancet Diabetes Endocrinol. 2015;3(1):35-42.

318. Holick MF. Vitamin D deficiency. The New England journal of medicine. 2007;357(3):266-81.

319. Zhao JG, Zeng XT, Wang J, Liu L. Association Between Calcium or Vitamin D Supplementation and Fracture Incidence in Community-Dwelling Older Adults: A Systematic Review and Meta-analysis. Jama. 2017;318(24):2466-82.

320. Silverwood RJ, Holmes MV, Dale CE, Lawlor DA, Whittaker JC, Smith GD, et al. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. International journal of epidemiology. 2014;43(6):1781-90.